DOCUMENT RESUME

ED 208 056 TM 810 812

AUTHOR	Bourke, S. F.; And Others
TITLE	Performance in Literacy and Numeracy: 1980.
INSTITUTION	Australian Education Council, Canberra.
REPORT NO	ISBN-0-642-06843-7
PUB DATE	81
NOTE	238p.
AVAILABLE FROM	Australian Government Publishing Service, P.O. Box 84, Canberra, 2600, Australia.
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	*Academic Achievement; *Comparative Analysis; Criterion Referenced Tests; *Educational Assessment; Elementary Secondary Education; Foreign Countries; *Mathematics Achievement; *Reading Achievement; Test Construction; Testing Programs; *Writing Skills
IDENTIFIERS	*Australia; Australian Studies in Student Performance

ABSTRACT

The Australian Studies in Student Performance project was a response by the Australian Education Council to community concern about Australian education standards. It was intended that the results of an empirical study in literacy and numeracy would make a useful contribution to the continuing debate about student achievement. Thus, a study was designed to allow statements to be made about student performance throughout Australia in 1980 on tasks in reading, writing, and numeration. Also, comparisons would be provided between 1975 and 1980 student performance. A rationale for the tests administered is given, as is a description of test development; information on the sampling design is provided, and the samples are described; and the administration of the testing program and preparation of data analysis are described. (Author/GK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED208056

Performance in Literacy and Numeracy: 1980

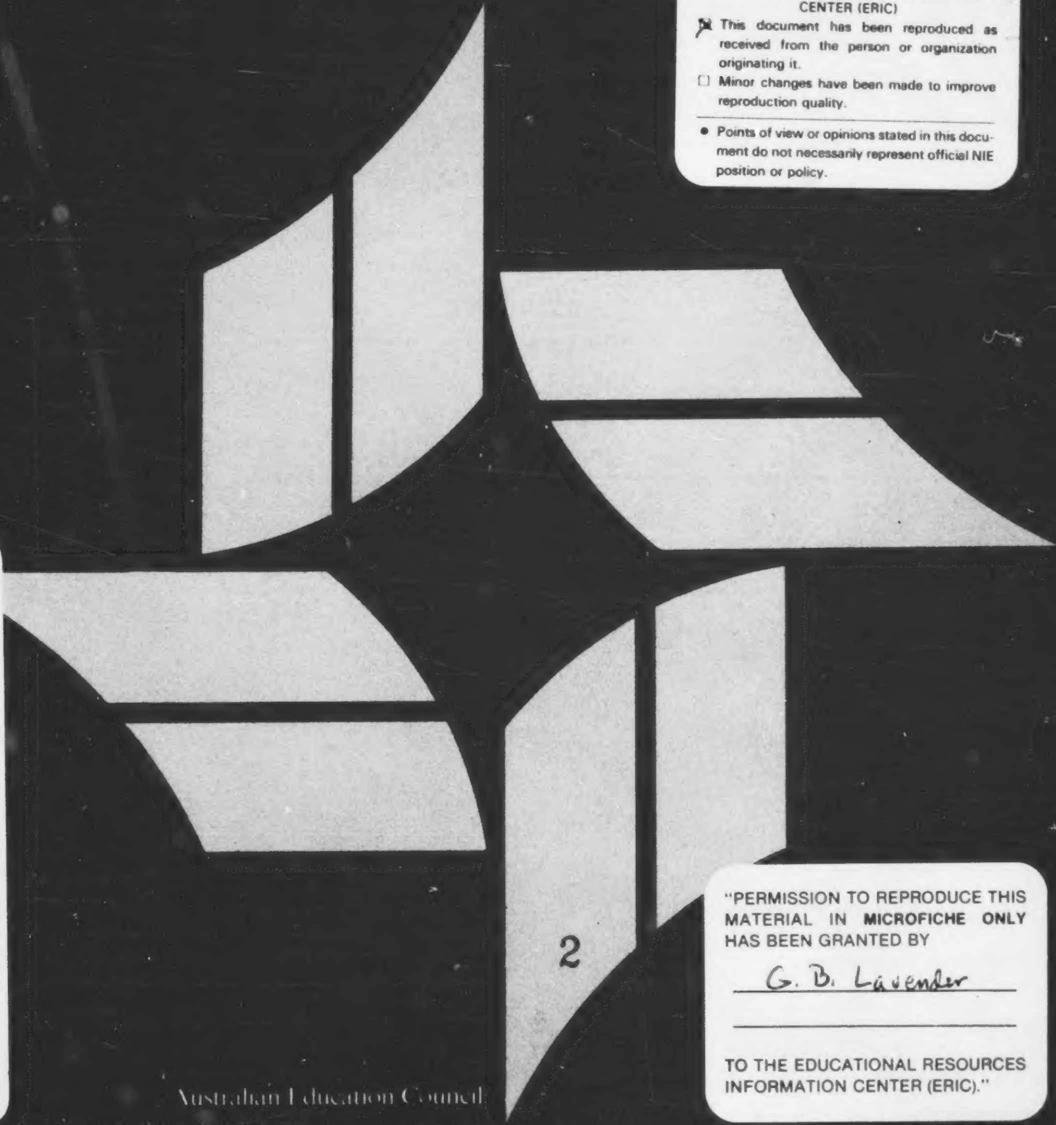
S. F. Bourke, J. M. Mills,
J. Stanyon, F. Holzer

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.



TM 810 812

"PERMISSION TO REPRODUCE THIS
MATERIAL IN **MICROFICHE ONLY**
HAS BEEN GRANTED BY

G. B. Lavender

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

Australian Education Council

Australian Education Council

Performance in Literacy and Numeracy: 1980

S.F. Bourke, J.M. Mills, J. Stanyon, F. Holzer

A Report to the Australian Education Council
on the Australian Studies in
Student Performance Project

Published by
the Australian Government Publishing Service
for the Australian Education Council
Canberra 1981

© Australian Education Council 1981

ISBN 0 642 06843 7

Requests to reproduce extracts to be addressed to the Secretariat,
Australian Education Council, 1st floor, 49 Spring Street, Melbourne 3000

Logo

Each quadrant of the AEC symbol depicts the silhouette of the capital letter E for Education. It symbolises the inter-governmental nature of the organisation and the eight Ministers who together constitute its membership and who are represented on the Council's working parties and task forces.

Printed by Hedges & Bell, Maryborough, Victoria

TABLE OF CONTENTS

		Page
Table of Contents		iii
List of Tables		vi
List of Figures		xi
Acknowledgements		xiii
CHAPTER ONE	INTRODUCTION	1
	Background to the Study	1
	Purpose and Outline of the Study	3
	Administration of the Study	4
	Structure of the Report	5
CHAPTER TWO	THE DEVELOPMENT OF THE ASSESSMENT INSTRUMENTS	7
	Introduction	7
	The Development of the Reading Tests	10
	The Development of the Writing Tests	17
	The Development of the Numeration Tests	24
	The Student Questionnaire	31
CHAPTER THREE	THE POPULATIONS AND THE SAMPLES	33
	The Populations	33
	The Sampling Design	35
	The Achieved Samples	36
	Estimation of Sampling Errors	38
	Possible Sample Bias Arising from Non-response	42
	Summary	44
CHAPTER FOUR	ADMINISTRATION OF THE TESTING PROGRAM	45
	Initial Contact with Schools and Follow-up	45
	School Administration and Return of Materials	46
	School Response Rate	46
	Reporting to Schools	48

		Page
CHAPTER FIVE	DATA PREPARATION	49
	Marking, Coding and Checking the Scripts	49
	Data Processing	50
	Calculation of Cutting Scores for Tests and Sub-tests	52
CHAPTER SIX	DESCRIPTION OF THE SAMPLES	59
	Sex of Students	59
	School Location	61
	Year Level of Students	62
	Age of Students	62
	Languages Other than English Spoken at Home	63
	Replacement Schools and Students	64
	Summary	65
CHAPTER SEVEN	PERFORMANCE ON THE READING TESTS	67
	The Dimensions of the 1980 Reading Assessment	67
	Procedures for Reporting Results	68
	Results for 10-Year-Old Students	69
	Results for 14-Year-Old Students	73
	A Comparison of Results on the Age Anchor Items	78
	A Comparison of Results 1975 and 1980	81
	Reading: differences Between Sub-Groups	83
	Summary	85
CHAPTER EIGHT	PERFORMANCE ON THE WRITING TESTS	89
	Student Performance on Criterion-referenced Tasks	90
	Student Performance on Globally-Scored Tasks	102
	Background Variables	109
	Summary	114

Page

CHAPTER NINE PERFORMANCE ON THE NUMERATION TESTS 117

Method of Reporting Used 117
Results for the 10-Year-Old Students 119
Results for the 14-Year-Old Students 127
Results by Background Variables 136
Comparisons of Student Performance in 1975 and 1980 142
Age Comparisons in Performance: 1980 146
Summary 149

CHAPTER TEN SUMMARY AND CONCLUSIONS 151

Test Development 151
The Student Samples 153
Performance in Reading 155
Performance in Writing 157
Performance in Numeration 158
Opportunities for Further Research 160
Final Comments 161

REFERENCES 163

APPENDICES 167

ONE Membership of Committees 167
TWO Letters to Schools and Parents 169
THREE Item Report 175
FOUR Examples of Students' Writing 181
FIVE Score Distributions and Reliabilities 219
SIX Project Costs 227

LIST OF TABLES

Page

Table 2.1 Classification of Items in Reading Test 10R by Form, Ability and Usage 12

Table 2.2 Classification of Items in Reading Test 14R by Form, Ability and Usage 12

Table 2.3 Link Items Between the 1975 and 1980 Reading Tests for 10-Year-Old and 14-Year-Old Students 14

Table 2.4 Common Items Between the 1980 Reading Tests for 10-Year-Old and 14-Year-Old Students 15

Table 2.5 Writing Objectives Represented at the 10- and 14-Year-Old Levels 20

Table 2.6 The Relationship Between Items and Objectives, Test by Test, and the Estimated Time in Minutes for Each Item 21

Table 2.7 Common Items in the Writing Tests, 1975 and 1980 (Link Items) and Between Age Levels in 1980 (Anchor Items) 23

Table 2.8 Classification of Items on the Numeration Test for 10-Year-Old Students 26

Table 2.9 Classification of Items on the Numeration Test for 14-Year-Old Students 26

Table 2.10 Link Items Between the 1975 and the 1980 Numeration Tests for 10-Year-Old and 14-Year-Old Students 30

Table 2.11 Common Items Between the 1980 Numeration Tests for 10-Year-Old and 14-Year-Old Students 31

Table 3.1 Australian Student Populations by Type of School and Age: August 1979 34

Table 3.2 Summary of Samples by State and Territory 37

Table 3.3 Weight Factors by State and Type of School for the Samples of 10-Year-Old and 14-Year-Old Students 39

Table 3.4 Approximate 95% Confidence Limits for Individual Items on the Reading, Numeration and Writing Tests for Overall Percentages: Australia 1980 40

Page

Table 3.5 The Magnitude of Four Standard Errors and Actual Percentage Differences Required for Significance for 1980 Comparisons Between Items for the Total Samples 41

Table 3.6 The Magnitude of Four Standard Errors and Actual Percentage Differences Required for Significance for 1975 and 1980 Comparisons in Performance for the Total Samples 41

Table 3.7 95% Confidence Limits for Reported Percentages by Sex and Percentage Differences Required for Significant Differences in Performance by Sex: 1980 42

Table 4.1 Response Rate of Schools Drawn in the Primary and Secondary School Samples 47

Table 5.1 Cutting Scores and Probability of Correct Classification of 10- and 14-Year-Old Students for the Reading Sub-tests and Total Tests 54

Table 5.2 Cutting Scores and Probability of Correct Classification of the 10- and 14-Year-Old Students for the Numeration Sub-tests and Total Tests 55

Table 5.3 Probabilities of Correct Classification of Students with Different Scores on the Reading Sub-test: Words in Context for 10-Year-Old Students 56

Table 6.1 Distributions of the 10- and 14-Year-Old Student Samples by Sex, Location, Year Level and Age (After Weighting) 60

Table 6.2 Percentage Sample Bias for Sex, School Location, Year Level and Languages Spoken Showing Potential Mean and Maximum Effects on Overall Australian Results for 10- and 14-Year-Old Students 62

Table 6.3 Distributions of the 1980 Weighted Samples by Languages Spoken at Home Compared with the 1975 Weighted Samples 63

Table 6.4 Replacement Schools and Students: Distributions of Replacement and Originally-Selected Samples of 10- and 14-Year-Olds (Weighted Data) 65

Table 7.1 Results on Reading Sub-tests for 10-Year-Old Students 70

Page

Table 7.2 Results on Reading Sub-tests for 14-Year-Old Students 74

Table 7.3 Reading Comprehension Performance of Students Aged 14 Years in Four English-Speaking Countries 76

Table 7.4 Results on Sub-tests Common to Reading Tests for 10-Year-Old and 14-Year-Old Students 79

Table 7.5 Performance on the Six Link Items, 1975-1980, by 10-Year-Old Students 80

Table 7.6 Performance on the Seven Link Items, 1975-1980, by 14-Year-Old Students 80

Table 7.7 Cumulative Percentages of Students Correct on Link Items, 1975-1980 82

Table 8.1 Comparison of Performance in Writing a Telephone Message by Students Aged 10 and 14 Years in Australian Schools in 1975 and 1980 91

Table 8.2 Performance in Writing a Personal Message by Students Aged 14 Years in 1980, in Australian Schools 92

Table 8.3 Performance in Writing Letters and Addressing Envelopes of 10- and 14-Year-Old Students in 1980 96

Table 8.4 Performance in Writing Letters by Students Aged 10 and 14 Years 98

Table 8.5 Performance on Completed Criterion-Referenced Tasks by Students Aged 10 and 14 Years in 1980 102

Table 8.6 Means and Standard Deviations of Global Scores, Number of Words and Spelling Errors per 100 Words of Writing Tasks of 10- and 14-Year-Old Students 105

Table 8.7 Categorization of Number of Words Written by 10- and 14-Year-Old Students on Globally Scored Writing Tasks 106

Table 8.8 Spelling Error Rate of 10- and 14-Year-Old Students on Globally Scored Writing Tasks 107

Table 8.9 Correlations of Global Scores, Number of Words and Spelling Error Rate for 10- and 14-Year-Old Students 108

Page

Table 8.10 Distribution of Correct Responses on Three Writing Tasks by Male and Female Students Aged 10 and 14 Years 110

Table 8.11 Distribution of Global Scores by Sex at the 10- and 14-Year-Old Levels 111

Table 8.12 Distribution of Correct Responses of 10- and 14-Year-Old Students on Three Writing Tasks by School Location 111

Table 8.13 Distribution of Correct Responses of 10-Year-Old Students on Four Tasks by Year Level 112

Table 8.14 Distribution of Correct Responses of 14-Year-Old Students on Four Tasks by Year Level 112

Table 8.15 Distribution of Correct Responses on Three Writing Tasks by Original Students and Replacement Students 113

Table 8.16 Distribution of Correct Responses on Three Writing Tasks by Original and Replacement Schools 114

Table 9.1 Numeration Test for 10-Year-Old Australian Students: Proportions of Successful Students on Each Sub-test and the Total Test 118

Table 9.2 Numeration Test for 14-Year-Old Australian Students: Proportions of Successful Students on Each Sub-test and the Total Test 128

Table 9.3 Percentages of 10-Year-Old Male and Female Students Responding Correctly to Items Exhibiting a Significant Difference in Performance Between the Sexes 137

Table 9.4 Proportions of 10-Year-Old and 14-Year-Old Male and Female Students Successful on the Numeration Sub-tests and Total Test 138

Table 9.5 Percentages of 14-Year-Old Male and Female Students Responding Correctly to Items Exhibiting a Significant Difference in Performance Between the Sexes 139

Table 9.6 Year Level of Students: Numbers and Proportions of Samples and Proportions of Successful Students on the Total Numeration Test 140

Table 9.7 Replacement Schools and Students: Numbers and Proportions of Successful Students on the Total Numeration Test 141

		Page
Table 9.8	Comparisons of the Percentages of Students Correct on Eight Basic Items Between 1975 and 1980	143
Table 9.9	Comparisons of Cumulative Percentages of Students Correct on Link Items Between 1975 and 1980	144
Table 9.10	Percentages of 10-Year-Old Students Responding Correctly to Link Items Between 1975 and 1980	145
Table 9.11	Percentages of 14-Year-Old Students Responding Correctly to Link Items Between 1975 and 1980	146
Table 9.12	Percentages of Students at Both Age Levels Responding Correctly to Items Common to Both Tests	147
Table 9.13	Cumulative Percentages of 10- and 14-Year-Old Students Correct on Age Anchor Items Sub-test	148

12

LIST OF FIGURES

		Page
Figure 2.1	Items Requiring Criterion-Referenced Marking - Test by Test	22
Figure 2.2	Items Requiring Global Scoring by Impression - Test by Test	23
Figure 2.3	Tasks or Objectives for the Development of Items in the Numeration Tests	29
Figure 8.1	Stimulus Material for Writing a Personal Message, 10-Year-Old Students	91
Figure 8.2	Stimulus Material for Providing a Comment and a Title, 10-Year-Old Students	94
Figure 8.3	Stimulus Material for Writing a Letter Requesting Information, 10- and 14-Year-Old Students	95
Figure 8.4	Stimulus Material in the Letter Offering to Help, 14-Year-Old Students	99
Figure 8.5	Part of the Stimulus Material for the Story, 10-Year-Old Students	103
Figure 8.6	Stimulus Material for the Story, 14-Year-Old Students	104
Figure 9.1	Item Requiring the Application of Whole Number Skills to Problem Solving	119
Figure 9.2	Two Items Involving Estimation of Metric Lengths	120
Figure 9.3	Money Item Requiring Subtraction with Amounts Less than One Dollar	121
Figure 9.4	Reading Time from Watch-faces With and Without Numerals	122
Figure 9.5	Two Items Involving the Recognition of Fractions	123
Figure 9.6	Items Testing Counting and Ordering	126
Figure 9.7	Proportions of Successful 10-Year-Old Students on the Numeration Sub-tests Compared with the Total Test	127

Page

Figure 9.8 Whole Number Item Involving Application of Subtraction with a Community-based Usage 129

Figure 9.9 Most Difficult Item on the Numeration Test for 14-Year-Old Students 130

Figure 9.10 Item Involving Measurement of Mass in a Community Context 130

Figure 9.11 Item Involving Differences Between Times Shown on a Circular Watch-face and a Digital Clock 131

Figure 9.12 Items Involving Knowledge of Place Value 132

Figure 9.13 Item Requiring Computation with Whole Numbers Involving Four Digits 133

Figure 9.14 Two Difficult Items in the Application Sub-test 134

Figure 9.15 Proportions of Successful 14-Year-Old Students on the Numeration Sub-tests Compared with the Total Test 135

ACKNOWLEDGEMENTS

In a large undertaking such as this study the contribution of many people involved in the planning and execution phases goes unnoticed by all but the project staff. Those involved in committees have an important role to play as do others who are informally consulted from time to time. We thank those who were generous with their time in this way.

We would particularly like to acknowledge the assistance given by many other members of the ACER staff during the course of this project. More specifically Dr Warren Jones, Mrs Yvonne Allen, Mrs Anne Lowry, Mr Craig Dowsett and Mr Jeff Clancy provided essential support either periodically or throughout the study. Mr Ray Smith of the Victorian Education Department drew many of the illustrations used in the tests. The provision of computing facilities by ICI Australia Limited and the prompt assistance given by staff of the Computer Operations Section, in particular Mr Doug Livingstone, were also most appreciated.

Most important of all we are grateful for the co-operation of the 10 576 students and the principals and staffs of the 476 schools throughout Australia that took part in the study, sometimes under conditions of considerable difficulty.

This report has been prepared by the Australian Council for Educational Research as a result of a research project commissioned by the Australian Education Council. The statements and conclusions in the report are the responsibility of the authors and of the Australian Council for Educational Research, and do not necessarily represent the views of the Australian Education Council or of the State, Territory and Federal Departments of Education.

CHAPTER ONE

INTRODUCTION

Background to the Study

The Australian Studies in Student Performance project was undertaken in 1979 by the Australian Council for Educational Research (ACER) at the request of the Australian Education Council (AEC). The study took the form of a survey testing program of 10-year-old and 14-year-old students in normal schooling throughout Australia to enable statements to be made about the performance of these students in the areas of reading, writing and numeration.

Although undoubtedly there were many factors involved, including general community expressions of concern, the AEC's decision in 1979 to proceed with the assessment of literacy and numeracy levels in Australia would have been influenced by two significant reports which had become available in the preceding months. One of the reports was that of the Commonwealth Government's Committee of Inquiry into Education and Training set up under the Chairmanship of Professor B.R. Williams which endorsed moves already being made and recommended further initiatives for the development of instruments to measure student outcomes (Williams, 1979 a, 1979 b). The second report resulted from a Study Group (under the Chairmanship of Mr A.H. Webster) established by the Education Research and Development Committee (ERDC) specifically to consider the need for a national assessment program in Australia. A draft report of this Study Group circulated during 1978. The final report, published in 1980, summarized the current range of assessments being made throughout Australia and emphasized the importance of providing assistance with assessment to schools and teachers as well as recommending periodic national monitoring of general literacy (Webster, 1980). An earlier report, prepared by the House of Representatives Select Committee on Specific Learning Difficulties and published in 1976, had also recommended that nation-wide surveys of literacy and numeracy be conducted at regular intervals (Cadman, 1976:23).

Subsequent to the commissioning of the present study by the AEC, the Myers Committee recommended that the federal government increase its support for national monitoring of literacy and numeracy and of other data collections in education (Myers, 1980:184). The AEC received yet more support from the Senate Standing Committee on Education and the Arts (Chairman, Senator G.S. Davidson) which, in recognizing some hostility towards national monitoring of literacy and numeracy, stated its belief in the desirability of an ongoing assessment to determine the effectiveness of a nationally accepted education program directed towards achieving the highest possible levels of literacy and numeracy. Further the committee was of the opinion that criterion-referenced national monitoring, impartially administered, would be of assistance to the teaching profession in the longer term (Davidson, 1981:32).

While stressing the importance of the school as the focal point for measures of student performance (Williams, 1979b:7; Webster, 1980:35) both the Williams Committee and the ERDC Study Group indicated that they saw a need for provision to be made for periodic testing on a national scale of the basic skills of literacy and numeracy (Webster, 1980:37; Williams, 1979a:128). Such testing programs would provide the schools with national student performance information which they could use as a basis for assessing the performance of their own students. The programs would also assure interested community groups that schools and policy makers accepted the importance of literacy and numeracy for all students.

Alternatively, the AEC's national monitoring program could be viewed as growing out of and extending assessment programs such as the Victorian Primary Schools Assessment Program (Renehan and Wilkes, 1973), the New South Wales Enquiry into Literacy and Numeracy in the Central Metropolitan Region of Sydney (NSW Department of Education, 1976), the periodic testing of literacy and numeracy in Tasmania (Hocking, 1976; Dodson, 1976; Broadby, 1977; Blazely, 1978; Broadby, 1978), and the Australian Studies in School Performance project conducted in 1975 by the ACER (Keeves and Bourke, 1976; Bourke and Lewis, 1976; Bourke and Keeves, 1977). The 1975 ACER study was undertaken at the request of the House of Representatives Select Committee on Specific Learning Difficulties (see Cadman, 1976) and was supported financially by the ERDC. Mr Race Mathews, who was the initial Chairman of the Select Committee, saw very clearly the contribution an Australia-wide survey into literacy and specific learning problems could make to the deliberations of the Select Committee and to Australian education. The development for the 1975 project of criterion-referenced tests which enabled an assessment of the proportions of students who could perform specific tasks, and the rejection of the use of normative tests which are designed to rank-order students, stemmed from initial discussions with the Chairman of the Select Committee in early 1975.

The ERDC Study Group made it clear that its recommendation to undertake a regular national assessment program was not inconsistent with its support for school level curriculum determination. The fundamental nature of literacy and numerical calculation for all areas of the curriculum was seen as minimizing the danger of curriculum conformity that the testing of more content-specific subject areas would face (Webster, 1980:33).

It should also be pointed out that the student performance measures identified by the Cadman, Williams and Webster Committees as being required by the schools went beyond the basic levels of literacy and numeracy. The Cadman Committee recommended that the range of competencies measured in the 1975 study by the ACER be extended in the short term to include 'mastery of language and oral skills' (Cadman, 1976:23). The Williams Committee recommended that:

> ... a range of performance levels to be expected of pupils of varying abilities ... (be specified) so that the horizon is not confined to the minima. (Williams, 1979b:7).

The ERDC Study Group recommended the development of 'more advanced literacy tests' (Webster, 1980:37). However such measures were not contemplated for the Australian Studies in Student Performance project in 1980.

Although the ERDC Study Group on national assessment and the Committee of Inquiry into Education and Training had widely different terms of reference and the members of both groups brought quite diverse perspectives to their work, they both considered, in some detail, the need for measures of student performance. It is of interest that their recommendations and their warnings on assessment of student learning were so consistently parallel.

Purpose and Outline of the Study

The Australian Studies in Student Performance project was a response by the AEC to community concern being expressed generally and through published reports about education standards in Australia. It was intended that the results of an empirical study in the areas of literacy and numeracy would make a useful contribution to the continuing debate about student achievement in Australian schools. To meet this intention the study was designed to enable statements to be made about the performance of students throughout Australia in 1980 on specific tasks in reading, writing and numeration. It would also provide for comparisons of student performance in Australia between 1975 and 1980 on a selected range of tasks in these areas.

It should be mentioned at this point that the survey testing program proceeded amid some controversy as to the purposes and desirability of gathering data on student performance in reading, writing and numeration.

In particular there was concerted opposition to the program mounted by some teacher unions and parent groups. However it is not within the province of this report to canvass further the arguments for and against the study that was carried out.

New tests were developed for the study and were used to assess the performance of the 10-year-old and 14-year-old students selected in the samples which were designed to include 1000 students per state and 500 students per territory. The total designed samples were thus 7000 students at each age level. Testing was carried out in October, 1980 and each school taking part in the study was offered a copy of the results for their school. Almost all the schools which took part in the study accepted this offer. Results for the total Australian samples in terms of the proportions of students successful on specific items and sub-tests were used as the basis of reporting together with results by sex and school location, which were included in the research brief.

Administration of the Study

When the AEC was considering the possibility of setting up this study, a committee comprising Mr D. Swan (Director-General of Education, NSW) as convenor, Mr S.S. Dunn (Chairman, ERDC) and Dr J.P. Keeves (Director, ACER) was set up to develop a proposal for the national assessment of basic skills. The AEC accepted the proposal prepared and set up a Steering Committee for the project consisting of the original three committee members plus a representative of each state and territory Department of Education (or equivalent) and the Federal Department of Education. Details of the membership of the Steering Committee and the other committees referred to in this section are recorded in Appendix One.

The Steering Committee was responsible for the overall supervision of the project. It also set up a small Technical Committee from its own members to assist in the development of the tests to be used. In addition to this Technical Committee, the ACER set up a Test Development Committee to provide specialist advice in each of the areas of reading, writing and numeration. These Committees consisted of practising teachers, teacher educators and curriculum experts who advised the ACER staff involved on the nature of the tests and assisted in the selection of items used in the final forms of the tests. Comments on the trial tests were sought from each of the education departments represented on the Steering Committee and these were taken into consideration by the Test Development Committees in advising on the final forms of the reading, writing and numeration tests.

Structure of the Report

A detailed rationale for each of the tests in reading, writing and numeration is given in Chapter Two of this report, together with a description of test development. Information on the sampling design, the achieved samples and estimates of sampling error is provided in Chapter Three and the samples are described in Chapter Six, using the background variables of sex, school location, year level and languages spoken at home to assess the representativeness of the samples. The administration of the testing program and the preparation of the data for analysis are described in Chapters Four and Five.

The performances of the 10-year-old and the 14-year-old students in reading, writing and numeration are reported in Chapters Seven, Eight and Nine respectively. These chapters, which constitute the backbone of the report, provide information on student performance in terms of the proportions of students able to succeed on individual items and on groups of items forming sub-tests concerned with specific tasks. Comparisons of performance between the 1975 Australian Studies in School Performance testing program (generally referred to as the 1975 ASSP program) and the present program (referred to as the 1980 ASSP program) are also made in these chapters. Chapter Ten provides summaries of the test development exercise, the sampling, and the reading, writing and numeration results, and busy readers may well find this chapter alone adequate for their purposes.

It is suggested that, when reading this report, persons with a more detailed interest in the assessment of student performance in literacy and numeracy should examine closely the test rationale statements in Chapter Two to decide for themselves whether they agree that the tasks identified are essential or at least important components of literacy and numeracy, that is that the tasks are important for students to be able to perform if they are not to be disadvantaged in their everyday lives or in their further progress through the school system. On the assumption that at least some of the tasks identified are thought to be important, the chapters of results could then be considered to decide whether the types of items used were reasonable tests of the tasks. Finally the reported levels of student performance should be considered. It would seem that only in this way can readers assess for themselves the results of this study.

CHAPTER TWO

THE DEVELOPMENT OF THE ASSESSMENT INSTRUMENTS

Introduction

Criterion-Referenced Testing

Although numerous definitions and interpretations of criterion-referenced measurement have been suggested, most seem to consider that the purpose of a criterion-referenced test is to determine the status of a student with respect to some well-defined objectives (see for example Millman, 1974). One of the most widely used definitions is that first suggested by Glaser and Nitko:

> A criterion-referenced test is one that is deliberately constructed so as to yield measurements that are directly interpretable in terms of specified performance standards (Glaser and Nitko, 1971:653).

Glaser and Nitko go on to state that the performance standards are usually specified by defining some domain of tasks that the student should be able to perform. Representative samples of tasks are taken and are used to make a statement about the performance of each student relative to that domain. The specified performance standards, or criterion levels of a test, are then often set at values which the examiner feels are minimum passing scores for the test.

The requirement of criterion-referenced testing is not to maximize the range of scores to enable discriminations to be made between students as is the case with normative testing. In fact comparisons between students on the tests are usually of no interest in a criterion-referenced test situation - it is the performance of each individual student compared with the criterion level which is the focus of concern. The argument for the use of criterion-referenced as opposed to normative testing in education is perhaps best summarized by Popham and Husek (1969) in their discussion of the basic distinction between the tests and of the issues of score variability, item construction, test analysis and interpretation of results, and by Kriewall (1972) who deals specifically with item difficulty, the shape of score distributions and test reliability. In particular the use of criterion-referenced measurement for instructional purposes is supported by Nitko (1970:14).

Criterion-referenced tests may be constructed to assess student performance on a very limited range of tasks or they may be related to a wide spectrum of tasks. According to Millman (1973, 1974) the domain of tasks may form a single dimension or may be along many dimensions, but in either case the structure of the domain must be precisely defined. If a test assesses performance on a number of dimensions within the cognitive domain, then the status of the student should be determined separately for each dimension. For the purpose of assisting a student to be able to perform a range of specific tasks, it is not very helpful to know that he/she has a total score on, for example, mathematics - what is needed is a series of measures of his/her performance on the specific tasks of interest. Consequently, in designing criterion-referenced tests it is first necessary to specify, as accurately as possible, the performances required of the student and to state clearly what performances are being assessed.

A criterion-referenced approach to testing was adopted in the 1975 ASSP study and was used again in 1980 in the reading, writing and numeration tests, the development of which is described later in the present chapter. In general, test developers repeated the procedures that were used in 1975. First test objectives were defined as clearly as possible. Next, tasks were devised to reflect these objectives and priorities were established. Then items were developed to assess performance on defined sub-tasks. Representative items were then selected which met two criteria - they had high face validity and they had appropriate levels of difficulty. Lastly, these items were field-tested in a sample of schools and a final selection of items was made.

Criteria for successful performance on each item were not formally specified. Rather, as in 1975, 'each selected item specified a required level of performance in operational terms, thereby implying criteria for satisfactory performance' (Bourke and Keeves, 1977:259). Items were grouped into sub-tests and the method used to summarize performance on each sub-test was to calculate the proportion of students who reached a pre-determined level of performance, as indicated by the proportion of items answered correctly. This level of performance was taken to be 80 per cent for each test and sub-test, that is students were expected to be correct 80 per cent of the time when answering from the population of items covered by a test or sub-test. This criterion level was chosen because it was considered appropriate for the kinds of tasks being assessed and also because it was the level used in the reading and numeration tests in the 1975 ASSP study. Comparisons between the two test administrations would therefore be possible. A rationale for the 80 per cent criterion level, or cutting score, is discussed in Chapter Five of this report.

In judging the validity of the tests administered in 1980, two points must be borne in mind. Fundamentally, the content validity (Cronbach, 1961: 106) of the tests depends upon the extent to which the test items represent faithfully the domain of tasks in literacy and numeracy which they were intended to measure. Most importantly, these tasks must be recognized as essential or at least important for students aged 10 and 14 years to have mastered if they are not to suffer disadvantage in their everyday lives or in their continuing progress through the school system.

Field Testing

Field testing was carried out in late April and early May, 1980 in primary and secondary schools. The selection of schools was such that the sample of students included both high-achieving and low-achieving students.

The procedure was that contact was made with school principals and a co-ordinator was asked to arrange for groups of 30 students to be released for a two-hour testing program. The four ACER test developers themselves administered the tests. The success of the field testing program was due to thorough preparation on the part of co-ordinators and the co-operation of classroom teachers.

There were two forms of the reading and numeration tests and four of the writing tests. At the 10-year-old age level approximately 280 students sat for the tests; at the 14-year-old age level, approximately 360 students were tested. The tests were longer than what was desired for the final tests in order to allow for the 'wastage' of unsuitable items. At each age level, the sample of students was divided into two sub-samples; students in each sub-sample sat for a different test but there were common items in each test. There were items common to the two age levels as well. Some items from the 1975 ASSP study were also repeated: the intention was that the 1975 items should establish to what extent the new items were equivalent in difficulty and discrimination to the 1975 items.

Responses were scored by the ACER clerical/technical pool and, in the case of writing, by experienced teachers, and results were returned promptly to the co-operating schools. Requests for clarification of data printouts were answered by telephone or, in several cases, by follow-up visits to schools.

After field-testing, copies of the tests were sent to all members of the AEC Steering Committee so that comment and criticism could be made by interested authorities from state education departments. Respondents included officers from the various state curriculum and research branches and committees of teachers convened for the occasion. Comment took the form of general statements, often accompanied by detailed annotations of separate items. Given the different curriculum policies of the states, it was predictable that many contradictory comments would be received and, in some test areas, it was difficult to find a consensus of opinion. Nevertheless, all comments were given serious attention, both by the test development committees, whose functions were described in Chapter One, and by the test developers themselves.

A rationale for the tests in each of the areas of reading, writing and numeration and a detailed description of test development are now given.

The Development of the Reading Tests

Reading in the context of literacy

Broadly defined, literacy consists of the ability to read and write. By extension, the skills of reading and writing are often linked with the skills of numeration to form a definition of functional literacy like the one proposed by Unesco in 1963:

> A person is literate when he has acquired the essential knowledge and skills which enable him to engage in all those activities in which literacy is required for effective functioning in his group and community, and whose attainment in reading, writing and arithmetic make it possible for him to continue to use these skills towards his own and the community's development and for active participation in the life of his country (In Keeves and Bourke, 1976:6).

This is the definition that guided the development of the survey tests of reading for the Australian Studies in Student Performance, 1980.

Clearly the reference is to an adult but the two emphases of the definition apply equally to the 10-year-old and 14-year-old student populations at which the testing program was directed. The first emphasis is one of 'effective functioning' in a person's group and community: in terms of a 10-year-old student, the implied functions are still largely school-based, but, in terms of a 14-year-old student who is free to leave school a year later, the implied functions have been increasingly influenced by the demands of the wider community. The second emphasis in the definition is on the continued development of the essential skills of literacy: a 10-year-old will build on them as he/she enters the last year(s) of the primary school and makes the transfer to secondary school, while a 14-year-old will build on them through continued formal study at the secondary or post-secondary levels or as an early or later entrant to the workforce.

The central component of literacy, as defined above, is reading, which Ryan (1975:14) defines as 'the acquisition of meaning from print, or it might be said to be thinking stimulated by print'. Smith (1978:105) speaks of reading as 'asking questions of a printed text'; again the implication is that the reader's purpose is to gain meaning.

Objectives in the teaching and testing of reading

There is evidence that the whole person is involved in the task of reading; it is therefore imperative to consider both the affective and cognitive domains in proposing objectives for the teaching and testing of reading. On the one hand, there is the objective proposed by the National Assessment of Educational Progress (1974) that students should demonstrate behaviour conducive to reading including, for example, demonstrating knowledge of their own reading ability. On the other hand, it would seem self-evident that students should be able to read a variety of materials with understanding. This very general objective in the cognitive domain has been further divided by Calfee (1978), after Miller (1973) to include three sub-objectives. The first concerns decoding and related word-attack skills, the second concerns comprehension *per se* and the third concerns application, that is, practical literacy, in the sense of 'a person's ability to use written and other kinds of information to realize practical goals' (Calfee, 1978: 427, 428).

The rationale for the 1980 reading tests

A test of approximately 30 minutes duration was required for each age level. Test development began with an examination of the rationale of the reading tests in the 1975 ASSP study (ACER, 1976a). The reading tasks in that program were classified in terms of three dimensions, the form within which reading material is set, the usage or purpose of written material presented for reading and the level of cognitive ability or skill involved in a reading task. These dimensions were accepted as a useful basis for the 1980 program. In the dimension of form, it was decided to use three areas: words in context, reference materials and continuous prose. Two other areas

Table 2.1 Classification of Items in Reading Test 10R by Form, Ability and Usage

READING ABILITY	USAGE	FORM OF WRITTEN MATERIALS			NUMBER OF ITEMS	
		Words in Context	Reference materials	Continuous prose		
Decoding skills	Classroom		*1*		1	1
	Community				-	
Comprehension	Classroom	*2-6,7-11*		*12-16*	15	18
	Community			*22,24,25*	3	
Application (reference skills)	Classroom		*17-21*		5	8
	Community		*23,26,27*		3	
Total number of items		10	9	8	27	

identified in 1975, graphic communication materials and written directions, were given a lower priority and were not used in the 1980 tests. In the dimension of usage, the two categories named 'classroom' and 'social' usage were still seen as useful distinctions, although it was decided that they might be better expressed as 'classroom-based' and 'community-based' reading materials, in that the main audience for the former consists of

Table 2.2 Classification of Items in Reading Test 14R by Form, Ability and Usage

READING ABILITY	USAGE	FORM OF WRITTEN MATERIALS			NUMBER OF ITEMS	
		Words in Context	Reference materials	Continuous prose		
Decoding skills	Classroom		*1*		1	1
	Community				-	
Comprehension	Classroom	*2-6,7-11*		*12-17*	16	24
	Community	*27*		*23,24,26, 31,32,33,35*	8	
Application (reference skills)	Classroom		*18-22*		5	10
	Community		*25,28,29, 30,34*		5	
Total number of items		11	11	13	35	

students and, for the latter, of the general public. Items were developed in both these categories. With respect to the dimension of cognitive ability or skill, some reservations were felt about dividing certain general abilities into a large number of discrete abilities, as was attempted in the 1975 ASSP study. Rather, the three ability areas identified by Calfee (1978) above were accepted as defining essential general tasks under which sub-tasks should then be identified with as much detail as seemed permitted by the present state of the literature in the testing of reading. In this respect, useful guidance in the measurement of reading comprehension has been provided by Spearritt (1977), whose research indicates three experimentally identifiable components, namely, knowledge of word meanings, sentence comprehension and apprehension of semantic context, a factor readily testable by means of passage-type comprehension exercises.

Table 2.1 shows the classification of items for the 10-year-old reading test (Test 10R) and Table 2.2 the classification of items for the 14-year-old reading test (Test 14R).

Item development

The Technical Sub-Committee of the Steering Committee of the Australian Education Council established priorities for the content of the reading tests. The shortness of the required tests was a real constraint on item development. Therefore, rather than give superficial treatment to important areas, it was decided that the reading tests should be deliberately selective. Consequently two assessment instruments under consideration were deferred: these were an Attitudes to Reading Scale (Teale and Lewis, 1978) and the IEA Word Knowledge test, which had been used as a measure of verbal ability in Australia in 1970, 1975 and 1978. It was decided that the reading tests would be limited to two main thrusts, namely comprehension and application, specifically the comprehension of sentences and passages and enquiry (or reference) skills. Thus the 1980 reading tests have a similar structure to the reading tests of the 1975 ASSP study. Several items from the 1975 study were repeated in the 1980 program, in order to permit a direct comparison of student performance on the two occasions. It was realized that performance on the common items might be affected by their being placed in a different order in the 1980 tests but their re-ordering was unavoidable and possible effects were not investigated. Table 2.3 shows that there were six link items for the 10R test and seven for the 14R test.

Table 2.3 Link Items Between the 1975 and 1980 Reading Tests for 10-Year-Old and 14-Year-Old Students

Item Numbers (10-Year-Old)		Item Numbers (14-Year-Old)	
1975	1980	1975	1980
1	1	1	1
2	6	2	6
4	3	4	3
18	23	23	33
20	26	24	31
26	24	25	32
		33	29

A test development committee consisting of educationists and practising teachers approved the suitability of all items, which were criterion-referenced. In other words, each item set a reading task of a kind considered important in providing evidence of a student's present status with respect to essential skills of reading. It was claimed that, at each age level, a student could reasonably be expected to complete each reading task satisfactorily. All items but the first (alphabet completion) were in multiple-choice response form.

Field-testing was carried out in two samples at each level. The two forms of the tests provided a range of items wide enough to meet the recommendation of the Technical Sub-Committee that the tests, especially the one for 10-year-olds, should not have a heavy emphasis on community-based usage, as in a newspaper. Some items from the 1975 ASSP program were repeated in the field-testing in order to establish equivalent difficulty levels: these included all the simple sentences, both prose passages and several newspaper items.

Common items at the two age levels

The final reading tests had the same format: each test consisted of six sections, the first two of which were identical and the remaining four of which were similar in purpose. Apart from the link items between the 1975 and the 1980 testing programs, the 1980 tests contained seventeen items common to the two age levels; these items are listed in Table 2.4.

The structure of the final tests

A description of the final tests is now given, section by section. All sections of the tests were introduced and timed by the test administrator;

Table 2.4 Common Items Between the 1980 Reading Tests for 10-Year-Old and 14-Year-Old Students

Test	Equivalent Item Numbers																
10R	1	2	3	4	5	6	7	8	9	10	11	17	18	19	20	22	27
14R	1	2	3	4	5	6	7	8	9	10	11	18	19	20	21	23	29

at each age level, practice questions were given for sections 2, 3 and 4. Section 1 (alphabet completion) came under the heading of word-attack skills. It is realized that a comprehensive assessment of such skills could only be made by means of an individual student's spoken responses to an examiner. However it was considered important to assess a student's knowledge of the roman alphabet, which is basic for literacy in Australian society, both for learning to read and for purposes of independent learning. The items in Section 5 (enquiry or reference skills) extended the assessment of this basic skill.

Section 2 (comprehension of simple sentences) involved single-clause sentences, the vocabulary and idiom of which were considered to be within the experience of 10-year-old and 14-year-old students. Each sentence had to be matched with four picture stories or line drawings, only one of which interpreted the meaning of the given sentence exactly. The non-verbal response mode was designed to be fair to all students. These picture stories were adapted from the unpublished tests of French as a Foreign Language used by the International Association for the Evaluation of Educational Achievement (IEA) in 1971: Carroll (1975) has reported the results of the cross-national testing program.

Section 3 consisted of the comprehension of compound/complex sentences. Items of this kind did not appear in the Australian Studies in School Performance (1975). This section was intended to test comprehension skills in sentences that were structurally more involved than the previous ones, but which were still representative, in both vocabulary and idiom, of sentences commonly found in the school experience of students at age 10 and age 14. The intention was that students should seek certainty of meaning in the words presented. This section was seen as a 'bridge' between the comprehension of simple sentences and the comprehension of continuous prose in the form of part of a classroom text, as in the fourth section, or in the form of newspaper articles, as in the final section of the test.

Section 4 tested, at each age level, the comprehension of a short prose passage typical of classroom or expository usage. Correct responses to these passages depended on a student's understanding of reasonably simple vocabulary and of the conjuctions, pronouns and adverbs that marked the development of the writer's thought. These passages were taken from the unpublished reading comprehension tests of English as a Mother Tongue, developed by the IEA in 1969: the results of the cross-national testing program held in 1970 have been reported by Thorndike (1973).

Section 5 (enquiry or reference skills) contained items designed to test a student's ability to use enquiry or reference skills related to reference material of the type found in classrooms and in school or community libraries. These items required the efficient application of reading skills to the task of gathering information: they included, in each test, selecting the appropriate volume of an encyclopaedia, using the contents and index pages of a book and scanning a page of an encyclopaedia for specific details. It is clear that Section 5 contains items that go beyond reading skills in decoding and comprehension. However, it is claimed that such items are consonant with the definition of functional literacy which informed the development of the reading tests.

Section 6, the final section of both tests, was based on a simulated newspaper. This stimulus material was devised as an example of the tabloid-style reading matter - this includes pamphlets, booklets, brochures, magazines, TV guides, handouts etc. - that has a largely adult audience yet, increasingly, is consulted by quite young students in school and is sometimes used by primary school teachers to develop reading skills (see Cheyney, 1971, and Nita Smith, 1978).

The functions of a newspaper are threefold: it provides information, it entertains and it provides services such as advertisements. The test items related particularly to the first and last functions: some required the comprehension of the literal and inferential meaning of articles that had been selected from Australian newspapers to interest 10- and 14-year-old students and others required the use of skills directly comparable with those assessed in the previous section, including scanning, and consulting an index, a time-table and a directory.

Nita Smith (1978:48) stresses the adult connotation that a newspaper has for a child who, if he/she can decode newspaper articles, has 'arrived' in society. The test developers decided it was essential that a 14-year-old student, who is approaching the school leaving age, should show competence in such an adult-oriented task. They also decided it was desirable that an upper-primary school student should provide evidence of his/her growing ability to cope with the flow of print outside the controlled situation of the classroom. These decisions resulted in a greater emphasis being placed on newspaper reading in the 14-year-old test than in the 10-year-old test. The former test included 13 newspaper items out of 35, the latter only six out of 27, which was half the number used in the corresponding section of the 1975 ASSP study. Every care was taken to ensure that the stimulus material had appropriate interest and difficulty levels and that the quality and size of print were adequate.

The complete tests, together with details of their development and administration instructions, have been published separately (Mills and Holzer, 1981).

The Development of the Writing Tests

Introduction

The writing tests were expected to provide estimates of performance in 1980 of 10- and 14-year-old students on a range of writing tasks. The aspects of writing performance to be considered were those which would be commensurate with the continuance of formal education beyond the most elementary level and those which would facilitate effective participation in the community. These two aspects were not considered to be mutually exclusive in objectives and tasks but rather matters of emphasis in the development of the writing tests.

In common with reading and numeration, the 1980 writing tests were expected to provide results which would enable comparisons to be made with the results of the 1975 ASSP study and between the two age levels. In conjunction with background information obtained elsewhere in the survey separate analyses were also expected to be undertaken of the writing skills of different groups of students, such as male/female and metropolitan/non-metropolitan. As a consequence of these expectations it was anticipated that the tests would be developed along similar lines to those of the 1975 ASSP study (ACER, 1976b).

Scrutiny of the Tests

In addition to normal ongoing scrutiny by the project group, objectives, tasks, items and procedures were scrutinized by a Technical Sub-Committee and a Writing Test Development Committee. The latter consisted of members of the wider professional community with a special interest and expertise in children's writing. This committee examined in detail the various aspects of the writing tests and made appropriate recommendations. A further contribution to the development of the tests came from the state, territory and federal education departments of Australia, officers from which commented on the tests at significant stages.

Identification of Objectives

A useful starting point for the development of the 1980 writing tests was considered to be the writing objectives of the 1975 ASSP study. These were examined and further refined with some variation in emphasis. The view of writing taken in the 1980 test development was that of a purposeful activity in which several systems of knowledge or skills are integrated in the responses which are made to writing demands.

In order that students can write effectively in response to these demands they must know how to use the writing system, so the kind of response they make requires a practical application and an integration of a number of skills rather than a demonstration of knowledge about written language, as for instance in a study of grammar or word usage, or a demonstration of competence in isolated skill systems such as handwriting, spelling, punctuation and capitalization. This emphasis does not mean that these kinds of knowledge are not considered important; for the purposes of the test development they are relevant in so far as they contribute to the successful completion of a writing task.

The writing demands that individuals have to meet in the course of their lives vary in kind, number and complexity. In the 1975 ASSP study two main areas from which demands could arise were delineated. One pertains to individual needs in the development of ideas, the structuring of experience and the expression of personal feelings and the other to the demands society might make at a personal, group or institutional level.

Within both of these areas a wide range of demands of differing levels of complexity could be proposed. The complexity of a task was seen to derive in part from the interaction between the function a task was intended to serve and the audience for which the response was intended. The shape and quality of a response was also seen to be influenced by the context in which a demand is made and the attributes an individual is able and willing to bring to bear upon a task. The linguistic resources of the writer such as knowledge of genre and conventions, handwriting skills, lexical and syntactic skills as well as cognitive development, motivation and perceptions will affect the response to a task as will the constraints, pressures or congeniality of the context.

The focus of the writing tests was determined by the general objectives of the 1980 study, which were concerned with the abilities of Australian students at certain ages to meet the writing demands they are likely to encounter at school and in the community. Whilst the capabilities of individuals are important in the shaping of a written utterance the emphasis of the test was not on how or why a written utterance occurs but on what is written: the emphasis was on performative not developmental aspects of writing. However in the selection of objectives at each age level consideration was given to factors likely to affect response such as interest, familiarity and developmental maturity.

The nature of the study was likely to exclude certain objectives which might be highly desirable in a study of children's writing. For instance self-motivated writing was automatically excluded. Usual channels which are available for the conception and incubation of ideas such as talking, listening, researching and reflecting were excluded. There was little opportunity for either extensive revision or for the development of extended thought.

Whilst the interest in, and familiarity with context were important considerations in the construction of specific items, it was recognized that the contexts for each task would not be appropriate for every student and further that the test situation might affect the writing produced in individual responses.

Selection of Tasks and the Development of Items

In the 1975 study priorities were assigned to a number of writing tasks which had been considered for inclusion in the study. The tasks were re-examined in relation to the relevance of possible audiences, functions and contexts for the particular age groups. As a result some tasks were omitted or subsumed under other tasks and some priorities were re-assigned but a range of objectives was maintained.

Developmental stages, interesting and familiar contexts, different kinds of audience and function were taken into account on the more specific level of construction of items, as was the degree of control over the writing expected of the student. For some tasks the nature of the function and audience provided limited opportunities for selection of strategies, for others there were more choices. For highly specific, limited and directed tasks criteria for successful completion of an item were established. Other tasks allowed more choice in strategy and in the levels of complexity and sophistication of a response such as those which call upon a writer to express

Table 2.5 Writing Objectives Represented at the 10- and 14-Year-Old Levels

Objectives	10-Year-Old Level	14-Year-Old Level
1. Write a story	x	x
2. Write about personal experience	x	
3. Write attitudes and opinions		x
4. Write instructions	x	x
5. Write a message	x	x
6. Take a message	x	x
7. Write a report	x	
8. Write a letter to obtain information	x	x
9. Write a letter to give information		x
10. Address an envelope	x	x
11. Write the main idea in a cartoon	x	
12. Write relevant supporting detail in a cartoon	x	

and structure personal feelings, ideas and experience. For these tasks, general criteria were established, but no specific value was assigned to any of these criteria. Some items were assessed according to both specific and general criteria.

In all items the language used conveyed the meaning intended in the simplest way.

The Final Form of the Tests

From a pool of items, approximately double the number of items needed for the final form of the tests were chosen to be field-tested so that parallel items and the reliability of new items over a range of objectives could be tested. Both the tests and administration instructions were based on the format of the 1975 ASSP tests. After the trial tests were administered, the results were analysed. The trial tests were also circulated to state, territory and federal education departments for comment.

The selection of items for the final form of the tests depended on the following factors: items which represented objectives of high priority according to the Writing Test Development Committee, items about which those consulted had shown the greatest warmth, and items which were judged suitable from an analysis of the results of the field tests. Whilst there

Table 2.6 The Relationship Between Items and Objectives, Test by Test, and the Estimated Time in Minutes for Each Item

Test 10W/A

Title of Item	Obj.	Time
Look at the picture story below	11,12	2
Write a message	5	3
Write a report	7	8
Write about yourself	2	17
Total time		30 mins.

Test 10W/B

Title of Item	Obj.	Time
Take a message	6	5
Write a letter for information and address an envelope	8 10	8
Write a story	1	17
Total time		30 mins.

Test 14W/A

Title of Item	Obj.	Time
Take a message	6	5
Write a letter and address an envelope	9 10	10
Write a story	1	20
Total time		35 mins.

Test 14W/B

Title of Item	Obj.	Time
Write a message	5	2
Write directions	4	5
Write a letter for information and fill in an envelope	8 10	10
Write your opinion	3	18
Total time		35 mins.

is no guarantee that all persons will be satisfied with all items, during the development of the tests comments and criticisms from as wide a range of professionals in the field as possible have been taken into account. Viewpoints expressed have been balanced with overall objectives and the results of the trial tests to produce the final form of the tests.

The duration of the tests was to be approximately thirty minutes. Within this period sufficient time was to be allowed to enable students to complete all items. In order to represent as many objectives as possible, whilst maintaining an acceptable error quotient, two alternative versions of the tests at each year level were developed, namely: 10W/A, 10W/B, 14W/A, and 14W/B. At both levels one of the versions was to be given to half the sample. Seven items were tested at the 14-year-old level representing eight objectives and seven at the 10-year-old level representing ten objectives (see Table 2.5). Table 2.6 shows the relationship between items and objectives and also includes the estimated time for each item.

Certain items or parts of items were included to allow direct comparisons to be made with the 1975 ASSP tests. At the 10-year-old level there were three repeated (link) items and at the 14-year-old level there were two link items. Two items (anchor items) were common to both year levels. Table 2.7 lists the link and anchor items used in the 1980 writing tests.

Scoring of the Tests

The method of scoring used for different items varied according to the levels of complexity and sophistication of the task. The two methods used were criterion-referenced scoring, for which a number of specific criteria were established, and global scoring, which included scoring by impression and scoring by primary trait. Detailed instructions were prepared to guide markers using each method and, for global scoring, a preparatory workshop was held to rank and discuss representative responses. It should be noted that the objectives of the tests did not include the assessment of students in any one system of knowledge or skill. On some items, however, the number of words and spelling errors have been recorded but no qualitative analysis of spelling errors has been undertaken.

Criterion-Referenced scoring. The items which were scored by this method are listed in Figure 2.1. A number of specific criteria were established and each of these criteria was marked on a correct/incorrect basis. For a student to be considered to have successfully responded to any particular item, correct responses had to be made to a number of the specified criteria. So two kinds of data were available: scores on specified criteria and scores on successful completion of an item.

10W/A

Task 1. Look at the picture story below.

Task 2. Write a message.

Task 3. Write a report.

10W/B

Task 1. Take a message.

Task 2. Write a letter for information and address an envelope.

14W/A

Task 1. Take a message.

Task 2. Write a letter & address an envelope.

14W/B

Task 1. Write a message.

Task 2. Write directions.

Task 3. Write a letter for information and address an envelope.

Figure 2.1 Items Requiring Criterion-Referenced Marking - Test by Test

Table 2.7 Common Items in the Writing Tests, 1975 and 1980 (Link Items), and Between Age Levels in 1980 (Anchor Items)

Type of Item	Item, by title	Age Level
Link	Look at the picture story below.	10
Link	Write a message	10
Link and anchor	Take a message	10,14
Link	Write directions	14
Anchor	Write a letter	10,14

Global scoring: by impression. Five items were marked according to the impressions of experienced teacher-markers (See Figure 2.2). In the 1975 ASSP study teacher-markers were asked to mark as they normally would. In 1980, while a mark was not allocated to a specific criterion, markers were required to bear in mind five general areas considered important in the assessment of a particular piece of writing. These areas, described below, were based in part on dimensions of quality exemplified by Hunter-Grundin (1979:92).

accuracy - the degree to which the language describes situations, events, relationships and reflections,

originality - the degree to which the writer moves 'beyond what is given',

fluency - the ease of production,

mastery of conventions - such as spelling, punctuation, capitalization, appropriate format etc.,

social cognition - the degree to which the writer takes into account his audience, or supposed audience.

Each item was assessed independently by two markers, using a scale of 1 to 5. The final score was derived by adding the two scores and, in cases

10-Year-Old Tests	14-Year-Old Tests
10W/A	14W/A
Task 3. Write a Report.	Task 3. Write a Story.
Task 4. Write about Yourself.	
10W/B	14W/B
Task 3. Write a Story.	Task 4. Write Your Opinion.

Figure 2.2. Items Requiring Global Scoring by Impression - Test by Test.

where the two scores were discrepant by more than one mark, the test developer acted as a third marker and made the final judgment.

Global scoring: by primary trait. Primary trait marking is different from impression marking in that it focuses on traits which are important for a desired effect. One item, the letter in test 14W/A, was marked by this method. For this item, social cognition was considered to be of focal importance. The score took into account the degree of awareness shown by students in the selection and presentation of information to a known but formal audience in a manner which showed cognizance of the conventions of letter writing. Marking procedures were similar to those for impression marking.

The complete tests, together with details of their development and administration instructions, have been published separately (Stanyon, 1981).

The Development of the Numeration Tests

Introduction

The requirement for the numeracy component of the Australian Studies in Student Performance program for 1980 was to produce tests that assessed the numeracy of 10-year-old and 14-year-old students throughout Australia. It was not intended that the tests should cover all, or even most, topics in the mathematics syllabuses of schools, but only those aspects related to tasks considered necessary to either school or community requirements for numeracy. The tests were to last no more than thirty minutes.

More specifically the tests were expected to:

- provide estimates of the performances in 1980 of 10- and 14-year-old students on a range of tasks related to numeracy,
- provide results for 1980 which would enable comparisons to be made between results obtained in 1975 and in 1980, and
- provide sufficient background information on the students tested to enable separate analyses to be undertaken for different groups of students such as male/female and metropolitan/non-metropolitan.

Thus the tests were designed to measure performance on specific tasks in the same way as the tests used for the 1975 ASSP study (ACER, 1976c).

Dimensions of the Tests

The tests prepared in 1975 were based on three dimensions considered appropriate for numeracy; content area, ability required and likely usage (Bourke and Keeves, 1977:62). In the final forms of the tests, three content areas at the 10-year-old level, four areas at the 14-year-old level, two abilities and two usages were assessed (ACER, 1976c:11). The dimensions had been refined from much wider ranges of content, ability and usage originally identified as important.

Using the 1975 dimensions as a basis, content areas, abilities and usages were again identified in 1980 and refined to produce similar but somewhat different dimensions for the final forms of the new tests. Each of the new dimensions is first described in full and the aspects actually tested are then identified.

Content Area. As previously stated, the test content was not intended to cover all aspects of mathematics taught in schools. However seven areas were identified as being important for numeracy and were used as the basis for structuring the content dimension of the tests. The areas are:

- Number and numeration concepts.
- Properties of numbers and operations.
- Arithmetic computation.
- Measurement.
- Probability and statistics.
- Simple geometric concepts.
- Business and consumer mathematics.

It was clear that not all of these areas could be covered adequately by one thirty-minute test in the 1980 testing program. Emphasis was given to arithmetic computation and to measurement as had been done in 1975. However there were also items which tested some aspects of the other five content areas.

Ability. Three levels of ability or cognitive skill were identified as important for tests of numeration: knowledge, computation and application. Knowledge refers to being aware of and understanding a variety of number concepts (such as place value) and properties of numbers (such as counting and ordering). Computation requires the carrying out of one or more arithmetic operations where the student is not required to select the appropriate operation but is told which operation is needed. Application involves problem solving where the student is required to comprehend the problem, select the appropriate arithmetic or other operation and apply that operation in solving the problem.

The three ability levels form a hierarchy in that competence in computation implies some knowledge of number concepts and properties of numbers, and application requires both knowledge and computation skills. However, as was the case in 1975 when slightly different abilities were used, a rigorously hierarchical relationship would not always apply between the abilities (ACER, 1976c:6). The computation and application items presented

40

Table 2.8 Classification of Items on the Numeration Test for 10-Year-Old Students by Content, Ability and Usage

Content	Usage	Ability			Nos. of Items	
		Knowledge	Computation	Application		
Counting and Ordering	Community	*6,7,8*			3	3
	Classroom				-	
Numbers (Whole & Fraction)	Community			*21,25,26*	3	14
	Classroom	*1,2,5,10, 11,12,13*	*16,17,18,19*		11	
Measurement -General	Community	*3,4*		*32*	3	6
	Classroom	*35*		*27,30*	3	
Measurement -Money	Community			*15,22,24, 28,29*	5	5
	Classroom				-	
Measurement -Time	Community	*9,14,23, 33,34*		*20,31*	7	7
	Classroom				-	
Nos. of Items		18	4	13		35

Table 2.9 Classification of Items on the Numeration Test for 14-Year-Old Students by Content, Ability and Usage

Content	Usage	Ability			Nos. of Items	
		Knowledge	Computation	Application		
Whole Numbers	Community			*18*	1	11
	Classroom	*1,8,9,10, 11*	*13,14,15, 16,35*		10	
Fractions & Decimals	Community			*23,26,30*	3	7
	Classroom	*2*	*19,20*	*6*	4	
Measurement -General	Community	*28*		*5,31*	3	7
	Classroom	*12*		*4,22,33*	4	
Measurement -Money	Community			*24,29,32,34*	4	5
	Classroom	*7*			1	
Measurement -Time	Community	*17,21*		*3,25,27*	5	5
	Classroom				-	
Nos. of Items		11	7	17		35

to students in these tests were unlikely to be unfamiliar to many students. Consequently they could be completed by using learned routines which bypassed otherwise-necessary abilities lower in the suggested hierarchy.

Usage. Two types of usage were identified; classroom and community. The usages in no way form a hierarchy and overlap to a considerable extent. Classroom-based usage is associated with specific competencies that are essential for further learning in the classroom but are not necessarily essential for other purposes. Community-based usage (referred to as 'social' usage in the 1975 testing program) is related to competencies considered necessary for everyday living in Australia. Of course many competencies are relevant for both classroom and community usages, and often it is the context in which a particular test item is placed that locates it as assessing in the classroom or the community area.

Final Forms of the Tests

A combination of content area and ability classifications was used in reducing the seven content areas identified to a smaller number of sub-tests in the final forms of the numeration tests. As already stated, emphasis was given to arithmetic computation and to measurement.

At the 10-year-old level, five content sub-tests were formed out of the seven content areas listed. The first two areas (number and numeration concepts, and properties of numbers and operations) were coalesced into one group of items assessing counting and ordering. The third area of arithmetic computation was included in a sub-test concerned with computation, largely with whole numbers, but including the recognition of place value and common and decimal fractions. Three areas of measurement were also covered. Two sub-tests dealt specifically with money and with the measurement of time, while the third sub-test was more general in covering estimation of metric measures, reading a measuring instrument and reading a graphical representation. The items making up these five sub-tests, the three ability sub-tests and the two usage sub-tests for the numeration test for 10-year-old students are shown in Table 2.8.

The test prepared for the 14-year-old students was similar in structure to that just described for the 10-year-old students. Again there were five content sub-tests made up from the seven areas listed above. However counting and ordering was not included at this age level, and the use of common and decimal fractions was considered worthy of a separate classification at the 14-year-old level. Consequently two content sub-tests were concerned with whole numbers (largely computation and including recognition of place value and estimating), and with fractions and decimals (again largely computation and including place value and estimating). The three measurement sub-tests were the same as those described for the 10-year-old students. The items making up the five content sub-tests, the three ability and the two usage sub-tests for the test for 14-year-old students are shown in Table 2.9.

Development of Test Items

The content areas and abilities described above were combined to form a range of tasks or objectives which are shown in Figure 2.3. These tasks were then expanded to provide a detailed list of competencies or more specific objectives which were then allocated priorities. Test items were then developed for those competencies awarded a priority of one. Although it was clear that not all competencies with a priority of one would be represented by an item or items in the final forms of the tests, it was from this pool of competencies that item development proceeded. Thirty-five items were used at both age levels.

Form of Item. In the main, an open-ended format in which the student had to write a response on a line provided was used. When application or computation is required, the most realistic task involves the student in constructing his own answer. However some tasks such as estimating or stating place value would be much more complex if a multiple-choice format were not used. It was recognized that constructed response items are more difficult to score reliably than multiple-choice items, and a schedule of acceptable responses had to be drawn up for each item. In brief, it was decided that any recognizable form of the correct answer would be accepted and spelling would not be a criterion of correctness.

Wording of Items. For any item, where a choice had to be made between using mathematically-precise language and reducing the amount of reading to a minimum, the latter course was followed as long as it was clear what the student had to do. This procedure had also been adopted in developing the tests used in 1975 (ACER, 1976c:9). In some cases information necessary when answering an item was shown on an accompanying diagram, and this practice also helped to reduce the reading requirement of the numeration

1 Counting
2 Associating numbers and numerals with words
3 Using the decimal system of notation
4 Recalling basic addition and multiplication tables
5 Whole Numbers - using the four basic arithmetic operations
6 Decimal system - using the four basic arithmetic operations
7 Common fractions
  - identifying common fractions
  - using four basic arithmetic operations
8 Percentage
  - relating percent concept to decimal notation
  - calculating using percentages
9 Calculating using simple ratio, or proportion
10 Knowing simple mathematical terms and basic symbols
11 Using spatial knowledge.
12 Knowing and calculating with metric units of measurement
13 Rounding
14 Knowing simple geometric shapes
15 Calculating quantitative characteristics of simple geometric objects
16 Reading instruments for measuring length
17 Reading common measuring instruments
18 Reading instruments for measuring time
19 Performing calculations with measurements of time
20 Reading values from simple tables, charts and graphs
21 Collecting and interpreting quantitative data and constructing simple graphs
22 Interpret simple charts
23 Comprehending simple ideas of probability and using these ideas in everyday situations
24 Reading and interpreting simple plans and maps
25 Using skills appropriate to living in the Australian community

Figure 2.3 Tasks or Objectives for the Development of Items in the Numeration Tests.

Table 2.10 Link Items Between the 1975 and the 1980 Numeration Tests for 10-Year-Old and 14-Year-Old Students

Item Numbers (10-Year-Old)		Item Numbers (14-Year-Old)	
1975	1980	1975	1980
3	2	1	4
4	3	2	3
8	8	4	6
9	35	6	5
14	7	8	21
19	21	9	22
22	22	14	12
25	25	20	26
27	24	27	28
29	9	28	34
30	26	29	24
31	32	31	33
33	20	33	35

tests. However, for some items, information on a diagram was repeated briefly in the text of the item where this was considered desirable.

Items Common to the 1980 and the 1975 Tests. It was considered desirable to use a number of items in 1980 that were identical with items in the 1975 tests. These items form a clear link between the two testing programs and permit a direct comparison of student performance on the two occasions. There were thirteen 1975/80 'link' items at each age level, and the item numbers are shown in Table 2.10. The possible effects of different ordering of common items between 1975 and 1980 were not investigated.

In addition to the identical items shown in Table 2.10, there were a number of other items used in 1980 that were very similar to 1975 items. In some cases a slightly different format was used to test precisely the same numeration task, whereas in other cases the same format was used to test a similar but slightly different task.

Common Items Between the Two Age Levels. There were twelve items common to the tests prepared for the 10-year-old and the 14-year-old students. These items, with numbers shown in Table 2.11, provide a direct comparison of student performance across the two age levels tested in 1980.

Table 2.11 Common Items Between the 1980 Numeration Tests for 10-Year-Old and 14-Year-Old Students

Test	Equivalent Item Numbers											
10-year-old	1	10	11	12	13	14	16	17	18	19	31	33
14-year-old	1	8	9	10	11	17	13	14	15	16	25	21

The complete tests, together with details of their development and administration instructions, have been published separately (Bourke, 1981).

The Student Questionnaire

It was decided to seek general information about each student in the sample by questionnaire, as in 1975. Test co-ordinators were informed of the importance of the students' responses being as accurate and as complete as possible. To this end, administrators were encouraged to help any student in difficulty and to check all responses.

The student questionnaire asked for indispensable details such as name, address and name of school. Date of birth was also asked for so that a check could be made of the range of ages between 10 years 0 months and 10 years 11 months and between 14 years 0 months and 14 years 11 months, respectively. Students were asked to state their sex as well, so that any differences in test performance according to sex could be compared with the differences disclosed by the 1975 test program. Five questions used in the 1975 study to provide an indication of migrancy and language usage in the home were repeated in 1980. These were:

> In which country were you born?
> In which country was your father born?
> In which country was your mother born?
> How many years have you lived in Australia?
> Please write down the names of the languages most often used in your home (one or two languages only were asked for and an example was given).

In the 1975 ASSP program an extensive questionnaire for teachers provided supplementary data on matters such as students' performance in remedial classes, race, physical characteristics etc. However, as it had been decided that the 1980 test administration should cause as little disruption to school activities as possible, supplementary data were not collected from teachers on this occasion.

CHAPTER THREE

THE POPULATIONS AND THE SAMPLES

The Populations

Two separate populations were the subjects of this study. The first population consisted of students aged 10 years as at 30 September 1980 who were attending normal schools in any of the six Australian States or two Territories. Thus students attending special schools for the deaf, the blind or the educationally sub-normal were excluded from the study because it was recognized that these young people would be unable to undertake the tests in their standard form and under group administration. Facilities were not available to implement the modified individual testing that would have been required if these students had been included. The 10-year-old student population was chosen because these students would be in the middle to upper primary school years where the basic skills of literacy and numeracy should have been acquired.

The second population consisted of students aged 14 years as at 30 September 1980 who were attending normal schools in any of the six Australian States or two Territories. This was a parallel sample to the sample of 10-year-olds and was chosen because the 14-year-old students were in the middle secondary school stage at a time immediately prior to the end of the compulsory phase of their schooling. The literacy and numeracy levels estimated for this population would approximate the maximum levels of achievement in these areas for those students who would leave school at the earliest possible age. The 10-year-old and the 14-year-old student populations described were equivalent to those used in the 1975 Australian Studies in School Performance project and described by Keeves and Bourke (1976:12-15). Sampling from the same age populations enabled comparisons in performance to be made between 1975 and 1980. Age rather than grade sampling had been used in 1975 to enable data to be aggregated meaningfully across the six states and two territories.

The numbers of students in each of the populations as defined for the eight states or territories are shown in Table 3.1. The reference date for these figures was 1 August 1979, the date of the last school census for which information was available in June 1980 when the samples were drawn. The differences between the numbers of students in the populations between August 1979 and October 1980 when the samples were tested would be small.

Table 3.1 Australian Student Populations by Type of School and Age: 1 August 1979[a]

State/Territory	10-Year-Old Students			14-Year-Old Students		
	Government	Non-Government	Total	Government	Non-Government	Total
Australian Capital Territory	3,128	1,199	4,327	2,641	1,206	3,847
New South Wales	68,114	18,295	86,409	62,658	18,875	81,533
Victoria	54,433	16,946	71,379	48,312	17,835	66,147
Queensland	32,585	7,644	40,229	28,484	9,398	37,882
South Australia	19,917	3,047	22,964	19,055	3,591	22,646
Western Australia	20,111	3,592	23,703	16,970	4,525	21,495
Tasmania	7,023	1,137	8,160	6,339	1,236	7,575
Northern Territory	2,161	373	2,534	1,407	165	1,572
Totals	207,472	52,233	259,705	185,866	56,831	242,697

[a] Source: Australian Bureau of Statistics (1979:32-35).

The populations of students in each state and territory were also structured by the type of school attended, that is by the numbers of students attending government and non-government schools. This was done to enable the samples drawn to be checked against the known population structure and provided necessary information when sample weighting was being designed. It can be observed in Table 3.1 that the population of 10-year-olds at the time of this study was almost 260,000 students across Australia, with approximately 80 per cent in government schools. The population of 14-year-olds across Australia was over 240,000 students with slightly more than three-quarters attending government schools.

The Sampling Design

As the major purpose of this study was to provide information on the performance of students in the two populations for Australia as a whole, it was essential that all students were equally represented in the final samples. However the need to recognize the eight educational administrative regions across Australia and thus to sample by state or territory meant that equally-sized samples were designed for the six states and samples of half that size were designed for the two territories. Two-stage stratified random sampling was employed in drawing the samples for each state and territory. It was decided to use the same sample design and size that had been used in the 1975 ASSP study (Keeves and Bourke, 1976:16-19). A major benefit resulting from this decision was that the value of the design effects for samples of this type had been calculated and used in the 1975 study and could be used in calculating sampling error for the present study. Consequently samples of 1000 students in each state and 500 students in each territory were drawn using the method described in the following paragraphs. The state and territory samples obtained would then require weighting to produce representative Australian samples of 10- and 14-year-old students. At the same time any imbalance in the samples resulting from differential response rates could be corrected.

In the first stage of the two-stage sampling procedure, schools were sampled with a probability proportional to the number of 10- or 14-year-olds attending (as appropriate for the two separate age samples) from a complete list of schools for each state or territory. Such a list or sampling frame is prepared and regularly updated by the ACER Survey Section. The schools are listed in postcode order with all government schools first, followed by non-government schools. The number of 10- or 14-year-old students enrolled is shown for each school. Forty schools in each state and 20 schools in each territory were selected in this way. When a school declined to take part in the study, the school immediately following it on the list was selected. In this way a school was replaced by one with the same or a similar postcode and thus in the same locality.

In the second stage of sampling, 25 students at each school were selected at random from a list of students of the appropriate age provided by the school. Replacement students were also selected to allow for students transferring out of the school or otherwise not being able to take part in the study. In cases where a small school with fewer than 25 students of the appropriate age was selected, all available students of that age at the school were tested. As the sampling design required that 25 students were included from each school selected if all types and sizes of schools were to be adequately represented in the samples, in these cases the number of students was made up to 25 by also selecting students from the next school or schools on the sampling frame. These students were then considered to form a 'pseudo-school' in which 25 students were tested. Schools next to each other on the sampling frame tend to be of similar size as well as in the same locality. In this way small schools were adequately represented in the samples and, particularly at the 10-year-old level, more than 40 actual schools in a state were involved in the study. Thus with 25 students at each of 40 schools within a state, 1000 students were selected for testing. The designed samples for Australia as a whole were therefore 7000 students at each age level when the two smaller samples for the territories were added.

The Achieved Samples

As stated above the designed samples for this study were to consist of 1000 students at 40 schools in each state and 500 students at 20 schools in each territory giving a total of 7000 students in the designed samples at each age level. There are always losses from designed samples, such losses resulting from a variety of events. The actual numbers and distribution of students for whom usable information was obtained in this study has been referred to as the achieved sample. The achieved samples for this study are shown in Table 3.2. It is clear from the table that the achieved sample for some states much more closely approximated the designed sample than was the case for other states. Overall, usable information was obtained from 78 per cent of the designed 10-year-old student sample and from 73 per cent of the designed 14-year-old student sample.

All schools in the designed samples were plotted on maps as a first step in attempting to determine whether losses from the achieved samples formed a recognizable pattern. A map of each capital city was used to plot the metropolitan schools and a state map used to plot the non-metropolitan

Table 3.2 Summary of Samples by State and Territory

State/Territory	10-Year-Olds Designed	10-Year-Olds Achieved	10-Year-Olds Achieved/Designed %	14-Year-Olds Designed	14-Year-Olds Achieved	14-Year-Olds Achieved/Designed %
ACT	500	310	62	500	172	34
NSW	1000	959	96	1000	958	96
Victoria	1000	752	75	1000	642	64
Queensland	1000	654	65	1000	565	57
South Australia	1000	952	95	1000	990	99
Western Australia	1000	956	96	1000	871	87
Tasmania	1000	460	46	1000	668	67
Northern Territory	500	430	86	500	237	47
Total	7000	5473	78	7000	5103	73

schools. Within metropolitan areas there was no pattern to the losses, and non-metropolitan schools were also well-spread across each state. However at the 10-year-old level there was an imbalance between the achieved samples of the metropolitan and non-metropolitan schools in some states. The results of this imbalance in school location can be observed in Table 6.1 which shows that 53 per cent of the 10-year-olds attended metropolitan schools. In comparison the Australian census information for 1976 indicated that approximately 59 per cent of the Australian population lived in these metropolitan centres (Australian Bureau of Statistics, 1976). This percentage was taken as the best available indicator of the proportion of the population of 10-year-olds living in metropolitan centres. The possible effects of the imbalance between the population and sample proportions are explored in Chapter Six.

Another potential source of bias in the achieved samples was evident. Government schools were less well-represented in the achieved samples for some states than they should have been when compared with the total numbers of students reported in Table 3.1 attending government and non-government schools. Losses from the designed samples were clearly lower for non-government schools. Whereas differences in student achievement in the 1975 ASSP study had been small for school location, they had been large for type of school attended. Consequently it was decided to weight the achieved samples for school and student losses to remove the imbalance between government and non-government schools as well as between states or territories but not to move to the complexity of weighting for school location as well. If the achieved samples had been of the order of 90 per cent as had been the case in

1975, the weighting procedure would have differed in that government and non-government schools would not have required separate weights. Otherwise the procedure would have been the same as that actually used in 1980.

The formula used in weighting the results to remove imbalances resulting from the disproportionate sampling between states and territories, and from the different response rates for government and non-government schools is shown below. There were 16 strata of students formed from two types of school for each of eight states or territories. The formula refers to these strata.

$$\text{Weight factor} = \frac{\text{No. in national achieved sample}}{\text{No. in national age population}} \times \frac{\text{No. in stratum population}}{\text{No. in achieved stratum sample}}$$

The weight factors calculated by this method have been shown in Table 3.3 for each state or territory and for government and non-government schools. The weighted results were those subsequently used in calculating student results throughout this report unless otherwise specified.

Estimation of Sampling Errors

For simple random samples, the standard error for proportions is estimated by calculating

$$se = \sqrt{p(1-p)/(n-1)}$$

where p is the sample proportion, and

n is the sample size

Kish (1965:46).

The actual sample design used was a stratified random sample which made allowances for stratification by state and territory and by type of school, for cluster selection of students within schools, and for weighting necessary because of the disproportionate sampling of students between states and territories. It has been demonstrated that the use of formulae based on simple random sampling seriously underestimates the magnitude of sampling error when complex sample designs are used (Ross, 1976).

For the 1975 ASSP study the Tukey-Quenouille Jackknife procedure was used to estimate the magnitude of sampling errors for a range of item difficulties (Keeves and Bourke, 1976:23). As the same sample design was used in this study, sampling errors calculated were based on those obtained in 1975 with appropriate corrections for the different sample sizes. The effective sample size was that calculated after allowing for weighting between the types of school and between the states and territories. The formula for

Table 3.3 Weight Factors by State and Type of School for the Samples of 10-Year-Old and 14-Year-Old Students

Weight Factors	10-Year-Old Sample		14-Year-Old Sample	
	Govt.	Non-govt.	Govt.	Non-govt.
ACT	0.532	0.136	2.414	0.170
NSW	1.891	1.928	1.795	1.772
Victoria	2.240	1.488	2.768	1.364
Queensland	1.513	0.805	1.901	0.790
South Australia	0.509	0.506	0.477	0.503
Western Australia	0.526	0.501	0.533	0.471
Tasmania	0.458	0.175	0.260	0.167
Northern Territory	0.129	0.101	0.158	0.069

standard error was manipulated so that the standard error for a range of proportions for items in the 1980 study could be calculated as:

$$se_2 = se_1 \cdot \sqrt{\frac{n_1-1}{n_2-1}}$$

se_2 = standard error for proportions: 1980

se_1 = standard error for proportions: 1975

n_2 = effective sample size: 1980

n_1 = effective sample size: 1975

The standard errors calculated were then used to determine confidence limits for individual items in 1980 and the actual percentage differences required for significance between 1980 items, and between 1975 and 1980 items.

Sampling Errors for Items: 1980

Sampling errors were calculated for items ranging in facility from 10 to 90 per cent. It can be observed in Table 3.4 that sampling errors for items having facilities of 10 and 90 per cent were a minimum and for items with a facility of 50 per cent were a maximum. The range of two standard errors has been shown in the table. As the majority of items had facility values in the area of 70 to 80 per cent, the most appropriate 95 per cent confidence limits ranged from 1.7 to 2.5 per cent for reading and numeration items and from 2.1 to 3.5 per cent for writing test items depending upon age level.

In comparison with the confidence limits calculated for the 1975 study, it will be noted that at both age levels the limits for the 1980 ASSP

Table 3.4 Approximate 95% Confidence Limits for Individual Items on the Reading, Numeration and Writing Tests for Overall Percentages: Australia 1980

Reported Percentages 1980[a]	Reading and Numeration		Writing	
	10-Year-Olds %	14-Year-Olds %	10-Year-Olds %	14-Year-Olds %
90 or 10	±1.2	±1.6	±1.6	±2.3
80 or 20	±1.7	±2.2	±2.1	±3.1
70 or 30	±1.9	±2.5	±2.4	±3.5
60 or 40	±2.0	±2.7	±2.6	±3.7
50	±2.2	±2.7	±2.6	±3.8

[a] Percentages for the 1975 program were reported in Keeves and Bourke (1976:23).

project were wider for the reading and the numeration tests, and narrower for the writing tests. Whereas the overall sample sizes were smaller in 1980, only two forms of the writing tests were used instead of the three forms used in 1975. This meant that the numbers of students attempting each of the writing tests was greater in 1980.

To determine whether differences in the proportions of correct responses to items were significant, a magnitude of two standard errors for each of the items was taken as necessary, that is the reported percentages were required to differ by four standard errors to be considered as significantly different. This procedure resulted in a very conservative test for differences but was considered necessary given the nature of the project where, because of the very large sample sizes, even small actual differences which are unlikely to be educationally important could otherwise be statistically significant and thus be treated as if they were important. The use of more formal significance tests was rejected in favour of a common sense approach as recommended by Hays (1963:299). Consequently for the 1980 test administration, items with facility values of 90 or 10 per cent for 10-year-old students would need to differ by 3 per cent to be significant for reading and numeration and by 4 per cent for writing. Similarly items with a facility of 50 per cent for 14-year-old students would need to differ by 6 per cent for reading and numeration and 8 per cent for writing. The actual amounts of four standard errors and the consequent difference required to the next whole percentage are shown in Table 3.5. A magnitude of four standard errors taken to the next whole percentage was required because item results were reported correct to the nearest whole percent. Again this approximation made the significance testing more stringent.

Table 3.5 The Magnitude of Four Standard Errors and Actual Percentage Differences Required for Significance for 1980 Comparisons Between Items for the Total Samples

Reported Percentages 1980	Reading and Numeration				Writing			
	10-Year-Olds		14-Year-Olds		10-Year-Olds		14-Year-Olds	
	4se %	Diff.Req. %	4se %	Diff.Req. %	4se %	Diff.Req. %	4se %	Diff.Req. %
90 or 10	2.4	3	3.2	4	3.2	4	4.6	5
80 or 20	3.4	4	4.4	5	4.2	5	6.2	7
70 or 30	3.8	4	5.0	5	4.8	5	7.0	7
60 or 40	4.0	4	5.4	6	5.2	6	7.4	8
50	4.4	5	5.4	6	5.2	6	7.6	8

Sampling Error for Comparisons Between 1975 and 1980

When comparisons were made between performances in 1975 and 1980, the magnitude of four standard errors was slightly less for reading and numeration items and slightly more for writing items. These differences resulted from the varying numbers of students who attempted the tests in 1975 and 1980. In many cases the differences required for significance were the same as those shown in Table 3.5, but for the sake of completeness, the actual amounts at each level are shown in Table 3.6. It will be noted that the differences in performance required for significance varied from 3 to 6 per cent for reading and numeration and from 4 to 9 per cent for writing. The required differences were higher for the 14-year-old students in part because this sample was smaller than the sample of 10-year-olds.

Table 3.6 The Magnitude of Four Standard Errors and Actual Percentage Differences Required for Significance for 1975 and 1980 Comparisons in Performance for the Total Samples

Reported Percentages 1975 and 1980	Reading and Numeration				Writing			
	10-Year-Olds		14-Year-Olds		10-Year-Olds		14-Year-Olds	
	4se %	Diff.Req. %	4se %	Diff.Req. %	4se %	Diff.Req. %	4se %	Diff.Req. %
90 or 10	2.3	3	3.0	3	3.8	4	5.3	6
80 or 20	3.2	4	4.1	5	4.9	5	7.1	8
70 or 30	3.6	4	4.7	5	5.6	6	8.1	9
60 or 40	3.8	4	5.1	6	6.0	6	8.6	9
50	4.1	5	5.1	6	6.0	6	8.8	9

Table 3.7 95% Confidence Limits for Reported Percentages by Sex and Percentage Differences Required for Significant Differences in Performance by Sex: 1980

Reported Percentages	Reading and Numeration %				Writing %			
	95% Confidence Levels			Diff.	95% Confidence Levels			Diff.
	Male	Female	Sum	Req'd	Male	Female	Sum	Req'd
10-Year-Olds								
90 and 10	1.7	1.9	3.6	4	2.1	2.3	4.4	5
70 and 30	2.6	2.9	5.5	6	3.2	3.6	6.8	7
14-Year-Olds								
90 and 10	2.2	2.5	4.7	5	3.1	3.6	6.7	7
70 and 30	3.5	3.9	7.4	8	4.7	5.4	10.1	11

Sampling Error for Comparisons Between the Sexes: 1980

It is clear that when the total samples were split in order to make comparisons between any identifiable groups of students sampling errors for the groups would be higher than for the total samples. As a number of performance comparisons were made between the sexes in this study, confidence levels and the percentage differences required for significance were calculated for male and female students. These are shown in Table 3.7 where it can be observed that the difference between items required for significance at the 95 per cent confidence level ranged from 4 to 8 per cent for reading and numeration and from 5 to 11 per cent for writing. The actual numbers of male and female students are shown in Chapter Six, Table 6.1.

Possible Sample Bias Arising from Non-response

As shown in Table 3.2 the response rates for this study varied considerably between the states and territories and in some cases were quite low. In particular, student response rates of less than 80 per cent were obtained for both age samples in Victoria, Queensland, Tasmania and the Australian Capital Territory and for the sample of 14-year-olds in the Northern Territory. All other response rates were high but the possibility that the administrative regions with lower response rates might not accurately represent the populations of 10- and 14-year-old students in those states and territories had to be considered.

The results obtained in this study for each state and territory were weighted according to the relative populations of 10-year-old and 14-year-old students to produce estimates of the proportions of all Australian students who were able to perform specific tasks related to literacy and numeracy. The lower response rates for Victoria and Queensland were of considerable concern and it was considered advisable to examine more closely their results for any inconsistencies which might arise from sample bias. To do this the results for these states were examined for changes in performance between 1975 and 1980 compared with changes in performance for the most populous state, New South Wales, which had a very high response rate. The examination suggested that for most items the effect of possible sample bias in Victoria and Queensland on the overall Australian results reported in this study was negligible for the reading tests at both age levels and for the numeration test at the 10-year-old level. The possible effect on the numeration test results at the 14-year-old level was slightly greater but did not reach 0.5 per cent and thus was unlikely to have altered by one per cent the proportions of successful students reported for most test items.

There were however a few test items for which student performance differed considerably between 1975 and 1980 in Victoria and Queensland but not in New South Wales. As it was considered that these items represented the maximum possible bias arising from losses from the samples in Victoria and Queensland, the possible effects on the overall Australian results were estimated for these items. It was evident that possible effects for reading and numeration at the 10-year-old level and for reading at the 14-year-old level were less than 0.5 per cent. The maximum possible effect resulting from losses from the Queensland sample of 14-year-olds was also considerably less than 0.5 per cent. However, in the numeration test for 14-year-olds there was one item (Item 35) for which the Victorian student results for 1980 were much higher than in 1975 whereas the results for New South Wales and the other states differed by a much smaller margin. Although no other item approached this level of difference, the maximum possible effect on the overall Australian results for this item was estimated to be 1.3 per cent. Of course it may be that the concepts and skills required in this type of item, which involved estimating the result of dividing a four-digit numeral by a two-digit numeral, had received a considerably greater emphasis in Victoria between 1975 and 1980 and a lesser emphasis in other states. In this case it would be probable that the observed differences between the states was real and not a result of sample bias. However the possibility exists that the sample losses from Victoria resulted in an inflation of more than one per cent in the proportion of Australian students reported as successful for this item.

There were two other items in this test where it was estimated that the possible effect of losses in Victoria on the Australian results were between 0.5 and one per cent. Thus on a few items it would seem that non-response from Victorian secondary schools may have affected the numeration test results for Australia as a whole by as much as one per cent.

After weighting, the maximum possible effect of sample bias on the reported performance of all Australian students arising from lower response rates in the less populous states and territories would be much smaller than those discussed for Victoria and Queensland. In particular for Tasmania, the Australian Capital Territory and the Northern Territory, the possible bias would be negligible in each case. Consequently it is considered that even if the effects of the lower response rates for these three regions were consistent and could be added, their sum would not have been large enough to have altered the overall Australian results by even one percent. However student performance was not consistently higher or lower in Tasmania, the Australian Capital Territory and the Northern Territory and the nett effect of sample losses from the combined three regions was negligible.

Summary

The student populations, the designed samples and the achieved samples for the study have been described in this chapter. The achieved samples of 10-year-olds and 14-year-olds consisted respectively of 5473 and 5103 students, which represented approximately 78 per cent and 73 per cent of the designed samples. The achieved samples were weighted according to the state and territory populations of 10- and 14-year-old students and the response rates for government and non-government schools within each state and territory. The weighted samples were used in calculating performance levels for Australian students.

The samples of students for the 1980 study followed the same design as that developed for the 1975 ASSP study. Consequently it was possible to use the 1975 sampling errors as a basis for calculating and reporting sampling errors for the 1980 data. Finally the issue of potential sample bias arising from the lower response rates for some states and territories was discussed. In most cases the potential effect of lower response rates on the 1980 Australian results was very small to negligible. The exception was that performance on a few numeration test items by the 14-year-old Australian students may have been inflated by one per cent.

CHAPTER FOUR

ADMINISTRATION OF THE TESTING PROGRAM

Initial Contact with Schools and Follow-up

A random sample of schools from each state and territory was drawn from the ACER sampling frame, which had been set up in 1975 and which has been progressively updated since then. Early in June 1980 a letter was sent to the principal of each school in the sample, requesting that the school should participate in the study. This letter was accompanied by a sheet of background information on the study, an acceptance/rejection form, a memo from the project director to the teacher appointed school co-ordinator for the testing program, a student name form and a reply-paid envelope (see Appendix Two for an example of the letter of request and a copy of the sheet of background information). The primary school co-ordinator was asked to provide names of students aged between 10 years 0 months and 10 years 11 months on 30 September 1980; the secondary school co-ordinator was asked to provide names of students aged between 14 years 0 months and 14 years 11 months on the same date. It was intended that the second stage of the two-stage stratified random sample, the selection of 25 students for testing and five reserves, should be carried out by the clerical/technical staff of the ACER. Where a school declined to take part, an attempt was made to replace it by a school of the same type and from the same geographic area as described in Chapter Three.

After a random sample of 25 students had been drawn, a list of names was returned to each school co-ordinator, together with a handbook for use in test administration. Subsequently, before the week for testing - 6 to 10 October 1980 - the test booklets were despatched to the co-ordinators, together with student questionnaires, detailed instructions for administration and labels for the return of the completed tests by freepost. Letters to parents were also included so that parents would have full knowledge of the testing program and be able to withdraw their child if they so desired. In that event, a reserve student would be substituted (See Appendix Two for an example of the letter to parents).

A follow-up was carried out by a reminder letter, despatched at the end of July, and by telephone contact, which was made on up to three different occasions. First, those schools that did not reply to the original and follow-up requests for participation were telephoned for a decision. Second, those schools that had indicated the possibility of their needing help with test administration were telephoned and informed that the ACER would pay for the services of a relief teacher where necessary. Third, well after the testing week, those schools from which no returns had been received were given a reminder call by telephone. In the last case it was found that certain schools had been forced to delay the tests for administrative reasons, others had withdrawn from the study and yet others, having administered the tests, were awaiting advice from teachers' organizations before deciding whether to return the completed tests or not.

School Administration and Return of Materials

The Handbook for School Co-ordinators contained instructions that would ensure the uniform administration of the tests. Tasks to be undertaken before the testing date included checking materials received, replacing a sample student with a reserve student where necessary, sending letters to parents of both sample and reserve students and preparing the test room. Tasks to be undertaken during testing included reading aloud the printed instructions accompanying the three tests and the questionnaire and carrying out the special procedures applying in the case of students incapable of completing a section of a test or a whole test. These procedures required that a balance should be struck between giving every student a chance to show what he/she could do and not causing him/her distress in a situation of almost total failure. Tasks to be undertaken after testing included checking that all tests were accounted for and that the information on questionnaires was complete and accurate and finally, returning all materials to the ACER as soon as possible.

As indicated, some schools found it necessary to engage a relief teacher for the test administration. In this case the school co-ordinator was asked to ensure that the relief teacher had studied the details of the Handbook for School Co-ordinators.

Almost all the completed tests were returned to the ACER within a month of the testing week. However for a number of reasons, including postal delays, tests were still being received in the closing weeks of the school year, and the last batch of tests arrived on Christmas Eve, 1980.

School Response Rate

As is usual in any large-scale testing program, some schools declined outright to participate. These schools were immediately replaced by schools with similar characteristics.

Table 4.1 Response Rate of Schools Drawn in the Primary and Secondary School Samples

State/ Territory	No. in Sample[a]	Refused/ Replaced[b]	Tests not Rec'd (not replaced)[c]	With- drawals[d]	Tests Rec'd[e]	% Rec'd
Primary						
ACT	23	1	1	9	13	57
NSW	46	1		2	44	96
Victoria	51	7	3	14	34	67
Queensland	51	2		16	35	69
SA	55	3	4	-	51	93
WA	50	5	2	4	44	88
Tasmania	47	1		25	22	47
NT	27	2		5	22	81
Totals	350	22	10	75	265	76
Secondary						
ACT	21	-		14	7	33
NSW	40	1		1	39	98
Victoria	40	5	2	11	27	68
Queensland	40	1	1	16	23	58
SA	41	2		-	41	100
WA	42	1		4	38	90
Tasmania	42	-		14	28	67
NT[e]	14	3	2	4	8	57
Totals	278	13	5	64	211	75

[a] The numbers of schools included exceeded 40 per State and 20 per Territory where small schools were selected and additional schools were needed to make up 25 students per school.

[b] Refusals shown here are those that would be expected in the normal course of events. The reasons given ranged from a planned school sportsday in the week 6-10 October, to some schools that have a policy of refusing all such requests. Refusals were replaced by the next school on the sampling frame.

[c] Schools in this category either were prevented from testing by an unforeseen event (e.g., widespread sickness of students and staff) or the completed tests were apparently lost in the mail.

[d] The withdrawals refer to schools which advised that, although they had agreed to participate in the program, they had withdrawn as a result of teacher union and/or parent association pressure.

[e] Insufficient secondary schools were available for 20 schools to be selected in the Northern Territory sample. The deficiency in student numbers was made up, in part, by testing more than one group of 25 students in one of the large secondary schools.

What was unusual about the 1980 ASSP program was that a number of schools, having agreed to participate, withdrew from the study so late that it was impossible to replace them. Written communications from schools, and telephone conversations with principals established that objections to the study on the part of teacher unions and parent organizations had influenced school councils, staff associations and principals in the decision to withdraw.

In a few instances, confusion arose between the ASSP program itself and the early investigations of the project commissioned by the Education Research and Development Committee to evaluate the impact of the ASSP program on schools. Unfortunately this confusion prejudiced some decision-makers in schools who had been undecided about participating and caused their withdrawal from the program.

Reference should be made to Table 4.1 for details of the response rate of schools drawn in the sample.

Reporting to Schools

It was planned that reporting would take place in two stages. First, participating schools that requested test results for their own students would receive them as soon as possible after the testing week. Then, following the completion of the report for the AEC, a summary of the overall results would be sent to the participating schools.

Test results were sent to schools progressively over a period extending from late in 1980 to early in 1981. Several recommendations were made for the interpretation of these results. It was recommended that principals and interested teachers should keep results confidential to their schools and should follow the school's normal practice if parents requested access to the results of their children. It was suggested also that the results might be useful in a general way in identifying the problems of individual students in the areas tested, but a caution was expressed against any attempt to obtain 'average' results for a school by totalling individual scores.

The summary report prepared at the same time as this report was also sent to participating schools.

CHAPTER FIVE

DATA PREPARATION

This chapter briefly describes the preparation of the raw data for analysis. The tasks involved were detailed and complex because of the sizes and nature of the two samples of students and the large number of achievement variables (items) to be recorded for each student. The fact that approximately half the students in each age sample did one form of the Writing Test while the other half did a different form contributed to the complexity of the data preparation phase of the study.

Marking, Coding and Checking the Scripts

Students had recorded their responses on the test booklets. In the reading tests, all items but the first were in multiple-choice response form. Most items in the numeration tests required constructed responses, the rest being in multiple-choice response form. Responses to the writing test items were necessarily in constructed response form.

All the tests were hand scored. The reading tests, numeration tests and the criterion-referenced items of the writing tests were marked by staff of the Clerical/technical section of the ACER. Experienced teachers marked those items in the writing tests that required impression marking and primary trait marking. In general, the marking principle followed ensured that any benefit of doubt was given to the students. This applied particularly to the constructed response items in the numeration tests and the criterion-referenced items in the writing tests: in both these instances markers accepted as correct answers that succeeded in communicating the required information in spite of poor handwriting and mistakes in spelling or other conventions. For example, when students were asked to write the time in the numeration tests, answers such as *20 too 5* and *40 passed 4* were accepted.

Test and questionnaire data were coded by the Clerical/technical section and punched on to cards for computing. The operations of scoring, coding and card-punching were checked systematically. Where a student had attempted a test but had not responded to an individual item, a missing data code of *eight* was recorded for that item. This code indicated that a student had elected not to answer the item for some reason, even though he/she had the opportunity to do so. This code was converted to the *No attempt* column used in Appendix Three where the results for every item on the reading and numeration tests are reported. A code of *eight* thus had a different meaning to the missing data code of *nine* (described later in this chapter) which indicated that a student did not attempt an entire test because he did not have the opportunity.

Data Processing

File Building and Cleaning

Separate files were built for each of the Reading, Writing and Numeration Tests and for the Questionnaire. Thus each student who completed all of the program had one card in each of four decks. When the two age samples are considered this meant that there were eight separate files built, four for each sample.

The OSIRIS III package mounted on the IBM 370 facility at ICI (Australia) was used for most of the data preparation and analysis in this study (Survey Research Center (SRC), 1973). Each file was ordered according to a seven-digit student identification number. This number was a composite of five separate variables - sample code (10- or 14-year-old), state/territory code, school number, whether it was a small school with fewer than 25 students, and student number within a school. The ordering was checked using the Osiris MERCHECK program (SRC, 1973:151-161). The apparently out-of-order cards were then checked to determine whether they were simply misplaced in the card deck or whether they had been given the wrong identification number. Hand checking was necessary because both types of error were found in each deck.

When the ordering of cards within a desk was correct a file was built using the Osiris FBUILD program (SRC, 1973:163-183). At this time a file dictionary was created which named every variable in the file. Responses to every variable were then checked for wild codes, that is values that should not have existed given the coding scheme used. For example, all reading test items except item 1 were multiple choice items with four alternative responses. Permissible values for the variables created for these items were one to four inclusive plus the missing data code *eight*. Any other value for these variables would constitute an error in the file which needed to be found and corrected. The Osiris WCC program was used to find the cases with errors for any variable and the TCOR program was used to correct these errors (SRC, 1973:209-220).

Test and Sub-test Scoring

It was necessary to create the overall test and sub-test scores that have been identified in Chapter Two for each of the three areas of reading, writing and numeration. The creation of test and sub-test scores was a two-stage process which was carried out for each file using the Osiris ICON program, both stages being completed in the same operation (SRC, 1973:317-348). The first stage involved the reduction of codes for each test item to two categories, correct (=1) or incorrect (=0), and these 'new' variables were added to the file. The second stage summed the number of correct responses for the appropriate items to give the test or sub-test scores which were also added to the file. The ICON RECODE facility was used to create the new scored variables and the ICON SUM facility to provide the test and sub-test scores.

As a check on these relatively large and complex scoring procedures, the answers of at least ten students were also re-coded, scored and summed by hand. Although this was a tedious and time-consuming exercise it was considered to be a necessary check on the accuracy of the procedures prepared for the mass scoring of results by computer. The separate files, once checked, were then able to be used for preliminary analyses of the data.

Merging the Files

The procedure adopted of creating separate files for each test and for the background data reduced the complexity of the data preparation to manageable proportions and also provided data that could be analysed sooner than would have been otherwise possible. However the data from the four files at each age level had to be merged so that the background data could be linked to the achievement data for each student. The merging also provided yet another check on the separate files that had been built.

The files were merged in pairs over three stages using the Osiris MMP program (SRC, 1973:253-264). This program provided the capacity for the creation of a merged file from the union of the two sets of data, and the reporting of mis-matches, that is where a case was present in one file and missing from the other. A missing data code of *nine* was used for each variable in the file from which a case was missing. A missing case indicated that the student concerned had not done that test, although he had taken part in at least one other part of the testing program. This was checked by returning to the student attendance sheet and, where necessary, to the original test forms for the relevant school. In some cases the student really had done the test and the appropriate information had to be substituted for the missing data codes on the merged file, again using the Osiris TCOR program. In other cases it was confirmed that the student had not attempted that particular test and therefore the missing data code used was appropriate.

Finally a weight variable was added to the merged file by means of the ICON RECODE facility. This variable was then available for use with any analyses involving data for the total Australian samples of students. The weights were those described in Chapter Three and shown in Table 3.3.

Calculation of Cutting Scores for Tests and Sub-Tests

In reporting the results of the 1975 ASSP project, it was decided for the reading and numeration results that student success in mastering a specific task would be indicated by the achievement of a mastery level of 80 per cent, that is, a student would be able to do the task 80 per cent of the time. (Bourke and Keeves, 1977:35-37). Consequently a mastery level of 80 per cent was set for each of the reading and numeration sub-tests and the total tests, and the appropriate cutting scores were calculated using a method developed by Morgan (1979). Basically the method involved two steps: first parameters were calculated for the total score distribution for each test, and secondly the total test score was taken as the prior distribution in a Bayesian framework using a two-parameter probabilistic model to determine the cutting scores which minimized student classification errors for each sub-test and the total test.

The same level of success and the same procedure in calculating cutting scores were adopted for the present study. In this way the results obtained would be comparable with the 1975 ASSP results. The 80 per cent mastery level also had the advantage of being high enough to make it very likely that a student who achieved this level really could do the task, yet it is not so high as to constitute an unreasonable requirement. Only a summary of the method used is presented here. The interested reader should consult Morgan (1979) for the complete methodology.

Parameters for the Total Test Score Distributions

In order to use the total test score distributions as prior information for the determination of cutting scores for the reading and numeration sub-tests and total tests, it was necessary to determine the beta distribution which best fitted the empirical distribution in each case. The values of the two parameters of the best-fitting beta distribution were determined using a fractile assessment procedure based upon the first three quartiles of the empirical distribution. The fit was then checked by comparing the mean, mode and other score points, particularly in the vicinity of the 80th percentile of the beta and the empirical distributions. The beta distributions used as Bayesian priors were:

Reading Test for 10-year-olds	$\beta(5.7, 1.0)$
Numeration Test for 10-year-olds	$\beta(6.2, 1.3)$
Reading Test for 14-year-olds	$\beta(9.6, 1.3)$
Numeration Test for 14-year-olds	$\beta(6.4, 0.9)$

The Method of Calculation

The cutting scores for each sub-test and total test were calculated by means of program BAYES, details of which were given by Morgan (1979:60-71). The program incorporates corrections for guessing and for carelessness in calculating cutting scores. The possibility of guessing the correct response to a test item depended upon the nature of the item. As both open-ended and multiple-choice items were used, the guessing parameter for each sub-test and total test had to be calculated separately. As can be observed in Table 5.1 the guessing parameter varied from zero for a sub-test consisting of all open-ended items, to 0.25 for a sub-test made up of all multiple-choice items with four alternatives. In the absence of any empirical means of determining an appropriate value for the carelessness parameter, this was taken to be 0.05 for each sub-test and total test as had been done in the 1975 ASSP study (Bourke and Keeves, 1977:282).

An appropriate loss ratio was also incorporated into the calculation of cutting scores. The loss ratio refers to the relative acceptability of the two types of possible error in setting a cutting score. These errors were the classification of a student whose true score was less than the criterion level as a successful or mastery student because his actual score exceeded the cutting score for a sub-test or test, and the reverse, the classification of a student whose true score exceeded the criterion level as not successful because his actual score failed to reach the cutting score. The loss ratio used was 1:1, that is the two types of errors of classification were considered to be of equal importance. The use of this loss ratio also minimized the sum of the two types of classification error.

The cutting scores calculated for each reading sub-test and total test are shown in Table 5.1 and the cutting scores for numeration are shown in Table 5.2. In each case the probability of making a correct classification at the cutting score is also shown. Given the loss ratio used, the probabilities of correct classification at the cutting score necessarily approximated 0.50 and were not permitted to drop below 0.40. Thus approximately half the students who scored precisely the cutting score would be misclassified. For short sub-tests this meant that relatively large numbers

Table 5.1 Cutting Scores and Probability of Correct Classification of 10- and 14-Year-Old Students for the Reading Sub-tests and Total Tests

Sub-test	Guessing		Careless	Test Length		Cutting Score		Prob. at Cut.Sc.	
	10 yrs	14 yrs	Both ages	10 yrs	14 yrs	10 yrs	14 yrs	10 yrs	14 yrs
Simple Sentences	0.25	0.25	0.05	5	5	4	3	0.57	0.43
Complex Sentences	0.25	0.25	0.05	5	5	4	3	0.57	0.43
Prose Comprehension	0.25	0.25	0.05	5	6	4	4	0.57	0.48
Ref. Skills (Class)	0.25	0.25	0.05	5	5	4	3	0.57	0.43
Ref. Skills (News)	-	0.25	0.05	-	5	-	3	-	0.43
Newspaper Comprehension	-	0.25	0.05	-	8	-	6	-	0.57
Words in Context	0.25	0.25	0.05	10	11	8	8	0.55	0.46
Reference Materials	0.22	0.23	0.05	9	11	7	8	0.53	0.49
Continuous Prose	0.25	0.25	0.05	8	13	6	10	0.45	0.54
Classroom Usage	0.24	0.24	0.05	21	22	17	17	0.55	0.48
Community Usage	0.25	0.25	0.05	6	13	5	10	0.62	0.54
Link Items	0.21	0.21	0.05	6	7	5	5	0.65	0.57
Age Anchor Items	0.24	0.24	0.05	17	17	13	13	0.40	0.50
Total Test	0.24	0.24	0.05	27	35	22	28	0.56	0.43

Table 5.2 Cutting Scores and Probability of Correct Classification of 10- and 14-Year-Old Students for the Numeration Sub-tests and Total Tests

Sub-test	Guessing		Careless	Test Length		Cutting Score		Prob. at Cut.Sc.	
	10 yrs	14 yrs	Both ages	10 yrs	14 yrs	10 yrs	14 yrs	10 yrs	14 yrs
Whole Numbers	-	0.04	0.05	-	11	-	8	-	0.61
Fractions & Decimals	-	0.06	0.05	-	7	-	5	-	0.62
Numbers (Whole & Fract)	0.05	-	0.05	14	-	10	-	0.45	-
Measurement-General	0.08	0.06	0.05	6	7	4	5	0.46	0.62
Measurement-Money	0	0	0.05	5	5	3	3	0.46	0.55
Measurement-Time	0	0.04	0.05	7	5	5	3	0.58	0.52
Knowledge	0.08	0.05	0.05	18	11	13	8	0.41	0.60
Computation	0	0.03	0.05	4	7	2	5	0.41	0.64
Application	0	0.05	0.05	13	17	9	12	0.46	0.46
Community	0.04	0.03	0.05	21	16	16	11	0.58	0.44
Classroom	0.05	0.06	0.05	14	19	10	14	0.45	0.54
Addition	0	0.05	0.05	5	4	3	2	0.46	0.45
Subtraction	0	0	0.05	5	8	3	5	0.46	0.48
Multiplication	0	0.05	0.05	4	5	2	3	0.40	0.51
Division	0	0.04	0.05	4	5	2	3	0.40	0.52
Link Items	0.04	0.08	0.05	13	13	9	9	0.42	0.45
Age Anchor Items	0.02	0.02	0.05	12	12	8	8	0 40	0.46
Total Test	0.04	0.04	0.05	35	35	26	26	0.45	0.50

Table 5.3 Probabilities of Correct Classification of Students with Different Scores on the Reading Sub-test: Words in Context for 10-Year-Old Students

Score	No. of Students	Posterior Prob. of Correct Classif.	Posterior B. Distribution
0	2	1.00	(5.7, 11.0)
1	2	1.00	(6.7, 10.0)
2	7	1.00	(7.7, 9.0)
3	7	1.00	(8.7, 8.0)
4	19	0.99	(9.7, 7.0)
5	42	0.95	(10.7, 6.0)
6	150	0.86	(11.7, 5.0)
7	292	0.69	(12.7, 4.0)
8*	615	0.55	(13.7, 3.0)
9	1365	0.78	(14.7, 2.0)
10	2952	0.92	(15.7, 1.0)
Total	5453	0.83	

* The cutting score was 8 for this sub-test.

of students could have been misclassified but for longer sub-tests smaller numbers were involved.

The posterior probabilities of correct classification for a given cutting score increase rapidly as observed scores move away from that point. This tendency may be seen in Table 5.3 where the posterior probabilities of correct classification for the range of scores for the reading sub-test, Words in Context, chosen to illustrate this fact, are shown. This sub-test consisted of 10 items, and at the cutting score of 8, the probability of correct classification was 0.55. The probability was 0.69 for 7 items correct and 0.78 for 9 items and, as can be observed in the table, the probability of the correct classification of students was much higher for other scores. This sub-test had a highly-skewed distribution, and almost 3000 of the 5453 students who had a score for this sub-test scored 10 out of 10. There was a 92 per cent probability that this large number of students had been correctly classified as mastery students for the sub-test. Overall when the probability of correct classification was applied to the number of students at each score point, it was found that approximately 83 per cent of the total number of students would have been correctly classified on the Words in Context sub-test.

The proportions of students who would have been correctly classified on other sub-tests according to the probabilities found for different test scores were also of the order of 80 to 90 per cent. Generally speaking this result was considered to be satisfactory given that some sub-tests were quite short. Overall for the total tests it was calculated that approximately 90 per cent of the students would have been correctly classified according to the stated criteria of guessing, carelessness and loss ratio used. More formally, test and sub-test reliabilities were estimated using the Marshall Beta index (Marshall, 1975). These are shown in Appendix Five.

CHAPTER SIX

DESCRIPTION OF THE SAMPLES

Some background information on each student in the sample was collected and is used in this chapter to describe the samples on which the achievement data in the following chapters were based. Weighted numbers and percentages of students are used in each case. As achievement comparisons have been made with the 1975 ASSP study, differences in the samples between the 1975 and the 1980 testing programs are also noted in this chapter and the potential effects of the differences on the overall Australian results are estimated. The potential effect is reported for each background variable in terms of the mean difference in performance for all items in each of the reading and numeration tests and for the maximum difference found for any individual item in each test.

It was indicated in Chapter Three that when a school or a student selected in a sample did not take part in the study, where possible another school or student replaced the originally-selected school or student. Some concern was expressed that these replacement schools and students might have different characteristics which could be linked with performance levels. Consequently whether a school or student had been selected in an original sample or was a replacement was also recorded and these distributions are also described in this chapter.

Sex of Students

There were slightly more male students in both the 10- and the 14-year-old student samples than would be expected. The reason for this small but consistent difference was not clear but was probably related to the proportions of single-sex schools in the achieved samples. The weighted numbers and percentages of the sample students by sex are shown in Table 6.1.

When compared with the 1975 ASSP study where the weighted numbers of male and female students were almost equal (Keeves and Bourke, 1976:97), the 1980 testing program involved almost two per cent more male students and two per cent fewer females. From the results obtained in the 1975 program where the female students had a higher level of performance than the male students, especially at the 10-year-old level (Bourke and Keeves, 1977:92-95), the higher proportion of male students in 1980 could be expected to depress the achievement of the total Australian sample to a small extent. Such a

Table 6.1 Distributions of the 10- and 14-Year-Old Student Samples by Sex, Location, Year Level and Age (After Weighting).

Students (Weighted)		10-Year-Old Sample		14-Year-Old Sample	
		N	%	N	%
Sex					
Male		2867	52	2661	52
Female		2605	48	2441	48
Location					
Metropolitan		2910	53	2946	58
Non-metropolitan		2562	47	2157	42
Year Level					
10 yrs.	14 yrs.				
3	7	4	-	37	1
4	8	1274	23	1341	26
5	9	3733	68	3228	63
6	10	455	8	436	9
Missing data		5	-	60	1
Age (Months)					
10 yrs.	14 yrs.				
120	168	410	8	360	7
121	169	439	8	445	9
122	170	475	9	431	8
123	171	459	8	472	9
124	172	512	9	446	9
125	173	492	9	436	9
126	174	495	9	412	8
127	175	400	7	418	8
128	176	457	8	463	9
129	177	452	8	398	8
130	178	405	7	362	7
131	179	470	9	415	8
Missing data		5	-	45	1
Totals		5472		5103	

possible depression could be referred to as a source of bias in the 1980 results.

Based on the differences found in the performances of male and female students for the reading and numeration items in 1975, it was calculated that the mean effect of the potential bias was negligible and the maximum possible effect was less than 0.5 per cent. This suggested that the higher proportions of male students in 1980 would have been unlikely to have affected the overall Australian results reported by one per cent after rounding. The percentage of sex bias in the samples and potential effects on the overall Australian results in reading and numeration are shown in Table 6.2. The signs in the table indicate the expected direction of any effect of the potential bias.

School Location

School location was defined as metropolitan if the school was located in any of the six state capital cities or in Canberra, and as non-metropolitan if located elsewhere. This was the same basis used for school location in the 1975 ASSP study (Keeves and Bourke, 1976:96).

There was a lower proportion of students (53 per cent) attending metropolitan schools in the sample of 10-year-olds than would have been expected from the 1976 Australian census information which indicated that approximately 59 per cent of Australians lived in the state capitals and Canberra (Australian Bureau of Statistics, 1976:2401-2408). However the proportion (58 per cent) was almost as expected for the sample of 14-year-olds. The numbers and proportions of metropolitan and non-metropolitan students at both age levels are shown in Table 6.1.

The discrepancy at the 10-year-old level resulted in part from a slightly higher participation rate in the study by primary schools in non-metropolitan areas. Insofar as there were some small differences in student achievement favouring metropolitan students found in the 1975 ASSP study (Bourke and Keeves, 1977:92-95), it might be anticipated that the higher proportions of students in the sample attending non-metropolitan schools in 1980 compared with expectations arising from the 1976 census data would tend to bias the 1980 results by depressing slightly the overall Australian results obtained by the 10-year-olds.

The potential mean and maximum effects of bias for school location on the overall Australian results have been shown in Table 6.2. It can be observed that the potential mean effects were negligible for reading and numeration at both age levels, and the potential maximum effects were negligible at the 14-year-old level. The potential maximum effects were higher at the 10-year-old level but did not reach 0.5 per cent for either reading or numeration.

Table 6.2 Percentage Sample Bias for Sex, School Location, Year Level and Languages Spoken Showing Potential Mean and Maximum Effects on Overall Australian Results for 10- and 14-Year-Old Students

Variable	10-Year-Old					14-Year-Old				
		Reading		Numeration			Reading		Numeration	
Per cent	Bias	Mean	Max	Mean	Max	Bias	Mean	Max	Mean	Max
Sex	-2	-0.1	-0.2	-0.1	-0.3	-2	-0.1	-0.2	-0.1	-0.3
School Location	-6	-0.1	-0.4	-0.1	-0.4	-1	0	-0.1	0	-0.1
Year Level	+3	0.3	0.6	0.3	0.8	+5	0.3	1.0	0.5	1.0
Languages Spoken	-1	0	-0.2	0	-0.1	-5	-0.3	-0.8	-0.3	-0.8
Nett Effect		0.1		0.1			-0.1		0.1	

Year Level of Students

The distributions of students by year level are also shown in Table 6.1 where it will be noted that approximately two-thirds of the students in each sample were at one year level, that is Year 5 for the 10-year-olds and Year 9 for the 14-year-old students. Approximately a quarter of the 10- and 14-year-old students were in Years 4 and 8 respectively while fewer than ten per cent were in Years 6 and 10.

These distributions were similar to those obtained for year level in the 1975 ASSP study where slightly more students were in Years 4 and 8 and correspondingly fewer in Years 5 and 9 (Bourke and Keeves, 1977:87). The slightly higher proportions of students in the higher year levels in 1980 could be expected to result in marginally higher levels of student performance on the basis of the link between year level and performance found in the 1975 study (Bourke and Keeves, 1977:92-95).

The potential mean and maximum effects of bias for year level arising from differences in the distributions of students between 1975 and 1980 are shown in Table 6.2. Whereas the potential mean effect for reading and numeration did not exceed 0.5 per cent at either age level, the potential maximum effect at the 10-year-old level was less than one per cent and at the 14-year-old level was equal to one per cent.

Age of Students

The age distributions of students are also shown in Table 6.1. As the samples of students were randomly selected from all students of the appropriate age, it would be expected that approximately equal proportions of students would

Table 6.3 Distributions of the 1980 Weighted Samples by Languages Spoken at Home Compared with the 1975 Weighted Samples

Language Spoken at Home	10-Year-Old 1980		14-Year-Old 1980		1975[a]	
	N	%	N	%	10 yrs %	14 yrs %
English only	4522	83	4204	82	85	87
English + other language	750	14	773	15	13	11
No English spoken	109	2	113	2	2	1
Missing data	92	2	13	-	-	-

a Adapted from Keeves and Bourke (1976:97).

be obtained for each month's difference in age. This proved to be the case both in 1975 and 1980 and no differences between the samples were present for this variable.

Languages Other than English Spoken at Home

The degree to which languages other than English were spoken at home is shown in Table 6.3. Information on this variable was collected to provide a check on the 1980 samples compared to the 1975 ASSP samples, the distributions for which are also shown in the table. Although more than four-fifths of the samples were made up of students from homes where only English was spoken, the proportion of 14-year-old students in this category was less in 1980 than in 1975. Consequently it might be expected that 14-year-old student performance in 1980 for Australia would be slightly lower than in 1975 where, generally speaking, more students from homes where only English was spoken had succeeded on the tasks tested compared with other students (Bourke and Keeves, 1977:94-5). If the missing data were eliminated from the calculations for the 10-year-old student samples, the proportions of students in each language category would be almost identical in 1975 and 1980.

The potential mean effect of bias on the overall Australian results arising from languages spoken at home was negligible at the 10-year-old level and less than 0.5 per cent at the 14-year-old level. The maximum possible effects of bias were also very small for both reading and numeration at the 10-year-old level. However the potential maximum effects at the 14-year-old level exceeded 0.5 per cent but were less than one per cent. These effects are also recorded in Table 6.2.

The concept of potential bias in the 1980 samples arose from a concern about losses from the samples. However there are other possible explanations for the higher proportion of 14-year-olds who stated that a language other than English was used in the home. Two possible reasons for the difference found in languages spoken at home between the 1975 and the 1980 samples of 14-year-old students suggest themselves. Both reasons are related to the increased awareness of and interest in the ethnic backgrounds of a high proportion of Australians and neither reasons implies a bias in the sample but simply reflects what may be a change in the population of 14-year-olds. On the one hand perhaps there was a greater tendency for languages other than English to be spoken in the homes of the 14-year-olds in 1980, although it might be expected that such a change would also be evident at the 10-year-old level. On the other hand, perhaps because of changing attitudes, some 14-year-olds were more willing to state in 1980 that a language other than English was spoken at home than their coevals had been in 1975. The data collected did not permit exploration of these or other possible reasons for the small differences found.

Replacement Schools and Students

Replacement schools and students were used in the samples under the conditions stated in Chapter Three. Replacement schools were always of the same type and in the same location as the school replaced. Replacement students were randomly selected from the same school in the same way that the students they replaced had been chosen.

It can be calculated from Table 4.1 in Chapter Four that approximately five per cent of schools in the samples were replacement schools. However as some schools did not have 25 students of the appropriate age and other schools were not able to test all students selected, the proportions of students from replacement schools might be expected to vary from the proportions of replacement schools. The numbers and proportions of students in the samples from replacement schools are shown in Table 6.4 where it can be observed that six per cent of the sample students at each age level were from replacement schools.

The numbers and proportions of replacement students within schools are also shown in Table 6.4. Whereas seven per cent of the 10-year-old student sample were replacements for other students, eleven per cent of the 14-year-olds were in this category. The higher proportion of replacement students at the 14-year-old level was probably related to the organizational pattern of most secondary schools where students came together in different groups

Table 6.4 Replacement Schools and Students: Distributions of Replacement and Originally-Selected Samples of 10- and 14-Year-Olds (Weighted Data)

Students (Weighted)	10-Year-Olds N	10-Year-Olds %	14-Year-Olds N	14-Year-Olds %
Schools				
Original	5150	94	4822	95
Replacement	323	6	281	6
Students				
Original	5068	93	4536	89
Replacement	404	7	567	11

for different subjects. These arrangements would make it more difficult to administer a testing program for students from several classes and three year levels. It was probably also true that the 14-year-olds had a higher rate of absenteeism than the 10-year-olds resulting in a need for more replacement students as was the case in 1975 (Keeves and Bourke, 1976:46).

In order to determine whether the replacement schools and students were different from the originally-selected schools and students in some systematic way which would affect the student performance levels obtained for Australia as a whole, the performances of the original samples of students had to be checked against the total samples. These results are reported in Chapters Seven, Eight and Nine for reading, writing and numeration respectively.

Summary

In this chapter the student samples were described and checked for non-representativeness compared with the samples drawn in 1975. On the basis of the links between the variables described in this chapter and student performance in 1975, it was possible to predict the effect on the 1980 results of any differences found between the 1975 and the 1980 samples. It was suggested that performance in 1980 would be slightly depressed by changes in the distributions of sex, school location (at the 10-year-old level only), and languages spoken at home (at the 14-year-old level), and slightly inflated by changes in year level. Age distributions were virtually identical on both occasions. Possible sample bias arising from differential response rates between states and territories was discussed in Chapter Three.

It would, of course, be possible to investigate the separate and combined effects of these variables on the performance levels of students by means of multivariate analytic techniques but these have not been attempted. Given the small differences involved, it was not considered that the time and money required for such an exercise could be justified in this project. On balance it was claimed that the cumulative effects of sex, school location, languages spoken at home and year level on overall Australian student performance would be small. The measures of potential bias for individual items indicated that there was virtually no effect on the overall Australian results arising from sex and school location differences in the 1975 and 1980 samples at either age level. There was the possibility of an inflation in results of up to one per cent because of year level differences, and, at the 14-year-old level, the possibility of a depression of almost one per cent in results because of differences in the distributions of languages spoken at home. There was almost no potential effect of languages spoken at home on the 10-year-old student results. However it should be emphasized that the bias effects reported are potential effects only and that there was no way of determining whether the potential was realized in the 1980 results. In any case it can be observed in Table 6.2 that the sum of the mean effects for these four variables was negligible as far as the overall Australian results were concerned.

Any effects on performance of the use of the relatively small numbers of replacement schools and students would also be expected to be small. Any differences in performance between originally-selected students and replacement students are reported in the results chapters.

CHAPTER SEVEN

PERFORMANCE ON THE READING TESTS

The Dimensions of the 1980 Reading Assessment

Tables 2.1 and 2.2 in this report indicate the way in which items in the 10-year-old and 14-year-old reading tests have been classified. Three general dimensions or categories of reading objectives were accepted by the test constructors, who then identified important tasks for assessment in each dimension and, finally, devised representative items to assess performance on these tasks or sub-tasks. Reading tests for school students have traditionally included items in the dimensions of student ability or skill and the form in which reading material is presented. The 1975 testing program added a third dimension, that of usage, or the purpose for which an author presents material to be read. The same three dimensions, ability, form and usage were used in the 1980 program as well. The range of reading tasks that might be treated in these three categories is so great that some explanation is due to clarify the scope of the 1980 reading tests and show the relationships that exist among the items.

The Ability Dimension

Three ability or skill areas were represented:

Decoding and related word attack skills
Comprehension
Application (reference skills)

The first area was represented by a single, but important item requiring knowledge of the alphabet. In effect, two thirds of each test concerned the comprehension of printed material, with an emphasis on material common to classroom learning. Reference skills were assessed on tasks that would be likely to be met both at school and in the community.

The Form Dimension

The choice of form in reading material may determine whether there is effective communication with a reader or not. Reading objectives in this dimension range from one-word instructions, supported by graphics, on food cartons to the sophisticated prose of school text-books. Three categories of form were distinguished for use in the 1980 reading tests:

Words in context and the relationships between words

Reference materials

Continuous prose

Assessment in the area of words in context received some emphasis, almost exclusively in relation to the classroom. Items based on reference materials naturally sought to assess a student's skill in applying general reading ability to problem-solving in learning with or without a teacher's direction, as indicated in the ability dimension. The comprehension of continuous prose included prose written expressly for a student and the prose of newspaper articles.

The Usage Dimension

Consideration was given to the four categories of purpose or usage which were identified in the construction of the 1975 reading tests: the categories were recreational, classroom or expository, utilitarian and personal. In 1975 the utilitarian and recreational categories were combined to form a category entitled 'social' usage and this was used as one of two categories, the other being classroom usage. This broad distinction was accepted for the 1980 tests but 'social' usage was renamed 'community' usage. Two different groups of readers are covered by these categories: the immature reader in the controlled conditions of the classroom and the mature reader in the adult community outside the school. It is clear that the categories overlap, in that a student is developing the flexibility to cope with reading tasks directed mainly at adults, but the distinction is a useful one for the assessment of a student's present status as a reader and his/her potential as a literate adult. Items representing classroom usage predominated at both age levels, especially at the 10-year-old level.

Procedures for Reporting Results

Results on the items used in the reading tests, in terms of percentages of students who were correct, who were incorrect and who made no attempt, are recorded in Appendix Three. A discussion of the results now follows: emphasis will be placed on groups of items, called sub-tests, within the three dimensions outlined above. Table 7.1 and Table 7.2 list these sub-tests according to dimension and present the percentages of students who achieved the mastery levels derived from the calculation of cutting scores as explained in Chapter Five. The effect of varying the mastery levels that were used can be seen by reference to the score distributions recorded in Appendix Five. In general, groups of fewer than five items were not seen to constitute meaningful sub-tests for reporting purposes but, where it is appropriate, comments will be made on such groups of items or on individual items. A total test score was also distinguished as a sub-test for students at each age level. In view of the wide range of items in the tests, this statistic had little specific meaning. However it was seen as a general guide to performance and a total test score was used both as collateral information in the calculation of cutting scores and as a criterion variable in reporting the performance of various student sub-groups. Sub-test and total test reliabilities have been reported in Appendix Five.

Results for 10-Year-Old Students

Results of sub-tests for these students are shown in Table 7.1.

Ability Dimension

Four sub-tests were distinguished in this dimension, three having to do with comprehension and one with reference skills.

To test the comprehension of simple sentences, five items were set. In each item a simple sentence was presented and a student was required to select from four accompanying line drawings the one that interpreted the meaning of the given sentence exactly. These five sentences, each one consisting of a single clause containing commonly used vocabulary, constituted the simplest reading comprehension exercise that could be devised in the form of words in context. It was clear that performance on these items would be a good indication of levels of reading ability. The five simple sentences were as follows:

Item 2	There are birds on the roof.
Item 3	He chose the biggest aeroplane.
Item 4	The children are swimming in the river.
Item 5	She is using scissors to cut the paper.
Item 6	The boy saw a fish in the fish bowl.

Performance on this sub-test was high. On an item-by-item basis an average of 98 per cent of 10-year-old students were correct. When considering performance on the sub-test as a whole it was found that a mastery level of four items correct was reached by 99 per cent of students. Out of 5473 students, there were 69 unsuccessful students in this sub-test; from the basic nature of the five reading items, it would seem that this group of students could not cope with most classroom reading tasks.

Five items were set to assess the comprehension of complex sentences. Students were required to read a sentence and choose from four statements

Table 7.1 Results on Reading Sub-Tests for 10-Year-Old Students

Dimension	No. of Items	No. of Items Required for Mastery	Successful Students %
Sub-Test			
Ability			
Comprehending Simple Sentences	5	4	99
Comprehending Complex Sentences	5	4	81
Comprehending a prose passage	5	4	70
Using Reference Skills (Classroom)	5	4	69
Form			
Words in Context	10	8	90
Reference Materials	9	7	77
Continuous Prose	8	6	70
Usage			
Classroom	21	17	74
Community (All newspaper items)	6	5	71
Total Test	27	22	71

the one that was true of the sentence. The first of these items is reproduced below.

Item 7 The girl was told that she could go if her father gave permission.

From this sentence we know

she is allowed to go.
she is not allowed to go.
she will not be allowed to go.
she might be allowed to go.

It is clear that these items were more demanding than the reading of simple sentences, containing as they did a wider range of vocabulary, tenses and idioms, and both coordinate and subordinate clauses. The average percentage correct on each item was 85 per cent and 81 per cent of students reached the required level of four items correct on the whole sub-test. Responses to the various alternatives showed that a number of students had found difficulty with a hypothetical condition in Item 10 and a distinction between all students learning about Japan and some students knowing a few words of Japanese, in Item 9. A total of 1046 out of 5473 students were unsuccessful on this sub-test.

As a natural progression from the comprehension of simple and complex sentences in isolation, a prose passage one paragraph in length was set for comprehension by 10-year-old students in the third sub-test. The paragraph, taken from the previously reported IEA study (Thorndike, 1973), provided five items in multiple-choice response form. Students had to read a description of an insect-eating plant, the sundew, and answer two questions involving the inferential meaning of the passage and three involving its literal meaning. An average of 77 per cent of students responded correctly to each item, and mastery, that is, four items correct, was achieved by 70 per cent of students. An analysis of alternative responses showed that the archaic use of the verb 'to light', meaning 'to come down', caused difficulty in Item 13 but not, apparently, in Item 14. The most difficult item proved to be the last one, involving a description of the sundew's leaves; the variety of adjectives used in the first few lines made a very careful reading necessary.

A sub-test of five items was designed to assess the performance of students learning independently by making use of the appropriate parts of reference books. These tasks were classified in the dimension of classroom usage. The first two items referred to a set of volumes in an encyclopaedia: first, students were required to name the volume in which information about Greenland could be found, and second, they were required to choose a volume which might have pictures of four different birds. The third item referred to the Contents page of a book about fish, and students had to locate a chapter likely to contain information about the great distances that fish travel. In the fourth item, the stimulus material was the Index page of an encyclopaedia and students had to locate the pages relating to travel by air. The fifth item concerned a page from an encyclopaedia, on which information about a famous person had to be traced. The first, third, fourth and fifth items were straightforward and an average of 82 per cent of students answered them correctly. The second item proved to be difficult as it required that students should derive the key word 'bird' from a question about kookaburras, hawks, eagles and crows and use the key word to locate the correct volume of an encyclopaedia; only 59 per cent of students were correct. However 69 per cent gained the mastery level of four items correct on the sub-test as a whole.

Ten-year-old students also took a sub-test of six items based on a newspaper: the items assessed both comprehension and reference skills. Results on this sub-test are discussed under the dimension of Usage (Community).

Form Dimension

Three sub-tests were identified in this dimension. Ten items tested words in context and the relationships between words, nine items tested the use of reference materials and eight tested the comprehension of continuous prose.

The first of these sub-tests consisted of the ten sentences read in isolation. A mastery level of eight items correct out of ten was required and 90 per cent of students attained this level.

The nine items testing the use of reference materials included the alphabet completion item, the five items associated with using reference skills in the classroom and the three items requiring the use of reference skills in newspaper reading. A performance level of seven items correct was required for this sub-test, and 77 per cent of students were successful.

The third sub-test in the dimension of form comprised the five items associated with reading a paragraph and the three items that required the comprehension of newspaper articles. A score of at least six items correct was required for mastery and 70 per cent of students were able to achieve that level of performance.

Usage Dimension

The first 21 items of the reading test, made up of the alphabet completion exercise, the comprehension of simple and complex sentences, the comprehension of a paragraph and the use of reference skills common to the classroom, formed the sub-test entitled Classroom Usage. In this sub-test, 74 per cent of students obtained the required score of 17 items correct out of 21.

The remaining six items in the test were all based on a newspaper and formed the Community Usage sub-test. For this assessment a simulated newspaper was prepared to represent the tabloid reading material that is such a feature of modern life. Ten-year-old students were given a tabloid sheet printed back and front and 14-year-old students were given a folded sheet of four pages. Questions with multiple-choice responses were contained in the reading test booklets. All articles in the newspaper had been culled from a range of recent Australian newspapers and they had been chosen for their interest value and their appropriate readability level. In several cases articles were re-set in order to increase the size or clarity of print. Three of the newspaper items assessed the comprehension of articles of various lengths and three assessed the use of reference skills directly comparable with those based on classroom usage. The three reference skills tested were: skimming to obtain an overall impression, reading a TV time-table to obtain information and using a directory to obtain information. Students performed well on this sub-test, except for the first item, which involved comprehension, and 71 per cent reached the required level of five items correct. This result was similar to the performance level demonstrated on the classroom reference skills sub-test and the same as the result on the total test.

There was little difference in the proportions of 10-year-old students who were successful on the sub-tests assessing classroom and community usage. However there were far more items, and a greater variety of items, in the assessment of classroom usage than in the assessment of community usage. When these facts are taken into account, it would seem fair to observe that approximately three quarters of the students did well on reading tasks familiar to them in school learning. At the same time, almost the same number of students seemed to have some familiarity with the style of reading found in newspapers.

Total Test Score

The total test score was also calculated as a sub-test. A score of 22 items correct out of 27 was required of 10-year-old students and 71 per cent were successful.

Results for 14-Year-Old Students

Results of sub-tests for these students are shown in Table 7.2.

Ability Dimension

Six sub-tests were distinguished on this dimension, four concerning comprehension and two concerning reference skills.

Fourteen-year-old students took the same five items as were used to test the 10-year-old students in the comprehension of simple sentences. In rounded figures, 100 per cent of students attained the mastery level of three items correct out of five. The actual result was that only three students out of 5103 failed to reach this level of performance.

The complex sentences sub-test was also common to both age levels. The level of performance was very high and 99 per cent of students gained the required level, which was three items correct. In this age group, Item 10 did not prove difficult but some 14-year-olds had the same kinds of difficulty with Item 9, concerning lessons about Japan, as the younger students. Sixty-two students were non-masters in this sub-test.

Table 7.2 Results on Reading Sub-Tests for 14-Year-Old Students

Dimension Sub-Test	No. of Items	No. of Items Required for Mastery	Successful Students %
Ability			
Comprehending Simple Sentences	5	3	100
Comprehending Complex Sentences	5	3	99
Comprehending a Prose Passage	6	4	83
Using Reference Skills (Classroom)	5	3	97
Using Reference Skills (Newspaper)	5	3	94
Comprehending Newspaper Articles	8	6	78
Form			
Words in Context	11	8	99
Reference Materials	11	8	93
Continuous Prose	13	10	72
Usage			
Classroom	22	17	94
Community (All newspaper items)	13	10	75
Total Test	35	28	83

It will be noted that, for 14-year-old students, a lower cutting score (3 items correct out of 5) was used in determining success in the comprehension of simple and complex sentences than for 10-year-old students (4 items correct out of 5). This apparent inconsistency between age levels on common items is explained by the different prior distributions of the two samples: the cutting score chosen minimizes errors arising from classifying students as masters or non-masters. In fact, the application of the more stringent cutting score would make no difference to the percentage of 14-year-old students gaining success in the comprehension of simple sentences and would lower by three per cent the percentage of successful students in the comprehension of complex sentences.

The 10-year-old sub-test comprising a paragraph to comprehend was parallelled at the 14-year-old level by a prose passage of three paragraphs about developments in the making of paper. Six questions were asked on the passage. The first two questions involved the literal meaning of the passage and the remaining four questions involved inferential meaning. An average of 79 per cent of students responded correctly to each item and the required level, that is four items correct, was achieved by 83 per cent of students. Item 16, which required students to identify the main point of a paragraph, caused extreme difficulty and the majority of students were incorrect: the different response rates for this item may be consulted in Appendix Three.

A cross-national comparison is of interest at this point. The prose passages used in the reading tests at both age levels were first used in the IEA Reading Project of 1970 (Thorndike, 1973). Thus information was available on the performance of 14-year-old students in England, New Zealand and the United States of America on the passage under discussion: Table 7.3 records the average percentages of students responding correctly to items in the 1980 ASSP study and in the 1970 IEA study in the three countries mentioned. The passage was a straightforward piece of expository prose and its theme gave no obvious cultural advantage to students of a particular country. The average percentages per item reported in Table 7.3 suggested that, in the passage in question, 14-year-old Australian students had performed better than their coevals in England and the USA and were marginally below New Zealand students. After correcting for the complex sample designs used in the 1970 and 1980 test administrations, the percentage differences recorded in Table 7.3 were tested for significance by applying formulae appropriate to uncorrelated proportions (Guilford and Fruchter, 1978:159, 160). Calculations indicated that, at the five per cent confidence level, there was no significant difference between the performance of 14-year-old Australian students in 1980 and 14-year-old New Zealand students in 1970. The difference in performance between the Australian students and the American students was not significant either. However there was a significant difference in performance between the Australian students and the English students.

The sub-test assessing reference skills applied to classroom tasks differed from the 10-year-old sub-test of the same name only in the fifth and last item, although a similar task, locating information on an encyclopaedia page, was assessed by that item. The average percentage of students correct on each item was 93 per cent and 97 per cent of students obtained at least three items correct, as required for mastery.

Table 7.3 Reading Comprehension Performance of Students Aged 14 Years in Four English-Speaking Countries

Country (year)	No. of Students	Average % Correct per item[a]
Australia (1980)	5103	79
England (1970)	3087	75
New Zealand (1970)	1955	82
United States (1970)	3503	76

[a] Data for England, New Zealand and the United States were retrieved from computer data tapes of results for the IEA Reading Project, 1970.

As has been explained in reporting results under the Usage dimension for 10-year-old students, the older students also answered items set on a simulated newspaper. The two pages presented to the 10-year-old students also appeared in the four-page newspaper for 14-year-olds, and this format allowed the setting of two common items. Thirteen newspaper items were taken by 14-year-old students; these items constituted two separate sub-tests, using reference skills and comprehending newspaper articles. It will be recalled that the six newspaper items for 10-year-old students constituted only one sub-test, that is the Usage (Community) Sub-test.

The three reference skills tested by means of the simulated newspaper were these: scanning to locate specific information, using a directory and using an index. There were two items on each of the first two skills and one on the third skill. The percentage of students attaining the required level of three items correct on this sub-test was 94 per cent. This result should be compared with the 97 per cent mastery level recorded above for reference skills applied to classroom tasks.

The last sub-test distinguished under the ability dimension was the comprehension of newspaper articles. This sub-test contained eight items and students were required to gain a score of at least six items correct; 78 per cent of students did so. It is interesting to compare this result with the mastery level of 83 per cent that was obtained in the sub-test on the comprehension of a prose passage typical of classroom reading.

Form Dimension

Three sub-tests were identified in this dimension. Eleven items tested words in context and the relationships between words, eleven items tested the use of reference materials and thirteen tested the comprehension of continuous prose.

The first of these sub-tests consisted of the ten sentences read in isolation together with the comprehension of a newspaper headline (Item 27). Ninety-nine per cent of students reached the required level of eight items correct out of eleven.

The eleven items which tested the use of reference materials included the alphabet completion item, the five items on reference skills (classroom usage) and the five items on reference skills (community usage). The required performance level of eight items correct out of eleven was achieved by 93 per cent of students.

The third sub-test in the form dimension consisted of the six items in the comprehension of a prose passage and the seven items based on the comprehension of continuous prose in the newspaper. In this sub-test students were required to answer 10 items correctly out of 13 and 72 per cent of them reached that level.

Usage Dimension

The first 22 items of the reading test, consisting of the alphabet completion exercise, the comprehension of simple and complex sentences, the comprehension of a prose passage and the use of reference skills applied to classroom tasks, formed the Classroom Usage sub-test. Ninety-four per cent of students reached the mastery level that was set by answering 17 out of the 22 items correctly.

The 13 items that remained in the test, all concerned with reading a newspaper, constituted the Community Usage sub-test. A score of at least 10 items correct out of 13 was required on this sub-test and 75 per cent of students reached this level.

Results on the two sub-tests in the dimension of usage show that a significantly higher proportion of students reached the mastery level on classroom usage than on community usage. However, in interpreting this difference, one should remember that the great majority of items in the classroom usage sub-test were common to both age groups. By contrast, only two of the 13 newspaper items appearing in the reading test for 14-year-old students were common to both age groups. In other words, three quarters of 14-year-old students demonstrated mastery in a sub-test (on newspaper items), which matched their expected levels of competence while over nine-tenths demonstrated mastery in a sub-test which more closely matched the expected level of competence of students four years their junior. The two results would appear to be similar.

Total Test Score

The total test score was also calculated as a sub-test. A score of 28 items correct out of 35 was required of 14-year-old students and 83 per cent were successful.

A Comparison of Results on the Age Anchor Items

A total of 17 items was common to the tests for 10-year-old and 14-year-old students. These "anchor" items have been identified in Table 2.4 of this report. Two sub-tests of five items each were included; they assessed the comprehension of simple sentences and the comprehension of complex sentences. To these must be added the alphabet completion item, four items assessing reference skills (classroom usage) and two newspaper items, one testing comprehension and the other testing skill in consulting a directory. The percentages of students who demonstrated mastery on the two sub-tests and the total anchor items have been recorded in Table 7.4.

As has already been noted, mastery was obtained by 99 per cent of 10-year-old students and, in rounded figures, 100 per cent of 14-year-olds in the comprehension of simple sentences. For the comprehension of complex sentences mastery was obtained by 81 per cent of 10-year-old students and 99 per cent of 14-year-old students.

The remaining anchor items were not treated as a sub-test, although it was of interest that an average of 76 per cent of 10-year-old students answered correctly each of the four common items on reference skills (classroom usage) as against 93 per cent of 14-year-old students. The alphabet completion item was answered correctly by 94 per cent of 10-year-old students and 98 per cent of 14-year-old students. The two common newspaper items were answered well by 14-year-old students, with an average of 92 per cent. Ten-year-old students had difficulty with the common newspaper comprehension item (Item 22 in their test), which concerned an aeroplane's destination, possibly because they were ignorant of the newspaper convention that requires the source of an article to be given prominence; only 55 per cent were correct here. However, their performance on the common reference skills item (Item 27 in their test, about a directory advertising pets for sale) was at the level of 82 per cent. It will be noted that 82 per cent of 10-year-old students were masters on the total anchor items, as against 98 per cent of 14-year-old students.

Table 7.4 Results on Sub-tests Common to Reading Tests for 10-Year-Old and 14-Year-Old Students

Sub-test	No. of Items	No of Items Required for Mastery		Successful Students %	
		10 y.o.	14 y.o.	10 y.o.	14 y.o.
Reading Simple Sentences	5	4	3	99	100
Reading Complex Sentences	5	4	3	81	99
Total Anchor Items	17	13	13	82	98

Special mention must be made of the performance of all students on the first item in the reading tests, which required students to write six letters which had been omitted from an alphabet printed in lower case. The letters were g, m, n, q, u and v. Any omissions or errors at all resulted in the item being marked as incorrect. A total of 299 10-year-old students and 124 14-year-old students were incorrect, including those who made no attempt. A random inspection showed that the most common error at both age levels was the inversion of the letters m and n and the next most common error was confusion over which letter followed the letter t (the letter v was a frequent choice). In an effort to find a possible explanation for a student's inability to complete the alphabet correctly, questionnaire data on several variables were consulted. Student and parental birthplace, the period a student had lived in Australia and the languages commonly spoken in the home were considered. It was revealed that, of the 299 10-year-old students who were incorrect on this item, 159 were born in Australia, had Australian-born parents and came from homes where only English was spoken; the corresponding figure for 14-year-old students was 55 out of 124. It would seem pertinent to observe, as was done as a result of the 1975 study (Keeves and Bourke, 1976:59) that students who have an incomplete knowledge of the alphabet could not confidently trace information through such modern aids to communication as telephone and street directories and classified advertisements in newspapers. If these deficiencies were not made good during the period of compulsory schooling, these students would be likely to suffer social and economic handicaps on leaving school.

In summary, the 14-year-old students were superior in reading ability to the 10-year-old students on the three common sub-tests, as judged by the proportions of successful students reported.

Table 7.5 Performance on the Six Link Items, 1975-1980, by 10-Year-Old Students

Item Number		Percentage Correct	
1975	1980	1975	1980
1	1	92	94
2	6	94	97[a]
4	3	96	98
18	23	91	88
20	26	94	92
26	24	80	86[a]
Mean percentage of students correct		91	93

[a] The differences between these proportions were significant in that they were greater than the sum of two standard errors for each of the 1975 and the 1980 test administrations.

Table 7.6 Performance on the Seven Link Items, 1975-1980, by 14-Year-Old Students

Item Number		Percentage Correct	
1975	1980	1975	1980
1	1	96	98
2	6	99	100
4	3	99	100
23	33	75	69[a]
24	31	85	86
25	32	79	78
33	29	89	96[a]
Mean percentage of students correct		89	90

[a] The differences between these proportions were significant in that they were greater than the sum of two standard errors for each of the 1975 and 1980 test administrations.

A Comparison of Results, 1975 and 1980

For each age level, there were several points of similarity between the reading tests administered in 1975 and those administered in 1980. The most obvious similarities appeared in the sub-tests assessing the comprehension of simple sentences and a prose passage and the sub-tests formed of items based on a simulated newspaper, assessing both comprehension and reference skills.

Some items were common to both test administrations: there were six such "link" items for 10-year-old students and seven for 14-year-old students. For 10-year-old students these items consisted of completing the alphabet, reading two simple sentences, skimming a newspaper page to locate an article, comprehending that article and locating information in a TV time-table displayed in the newspaper. For 14-year-old students the items consisted of completing the alphabet, reading the same two simple sentences as the 10-year-old students, using a directory in a newspaper and comprehending newspaper articles. Results on the individual link items have been recorded in Table 7.5 for 10-year-old students and in Table 7.6 for 14-year-old students.

In comparing performance on individual items in the tests for 10-year-old students it must be remembered that a four per cent difference was required for significance when item facilities were approximately 80 per cent and a three per cent difference when item facilities were of the order of 90 per cent (Table 3.6). On these statistics, differences between performances in 1975 and 1980 were significant on only two items, but in both cases they were in favour of the 1980 students. In the case of the tests for 14-year-old students, a six per cent difference was required for significance when item facilities were 60 per cent and a three per cent difference when item facilities were 90 per cent. These statistics indicate that the proportions of 14-year-old students answering the link items correctly in 1980 differed significantly from the 1975 study in only two cases: performance was lower in 1980 on a newspaper comprehension article and higher on the use of a directory in a newspaper. With regard to the last mentioned item (Number 29) it should be noted that the stimulus material had been enlarged photographically to fill a space one and a quarter times that of the original item from the 1975 testing program. This was done to provide an adequate size of print for the 10-year-old students to read as well.

Table 7.7 Cumulative Percentages of Students Correct on Link Items, 1975-1980

No. of Items Correct	10-Year-Olds		14-Year-Olds	
	1975 Cum. %	1980 Cum. %	1975 Cum. %	1980 Cum. %
	(six items)		(seven items)	
0	0.1	0.1	0.1	0.0
1	0.5	0.3	0.2	0.1
2	1.6	1.5	0.5	0.4
3	4.0	4.5	2.0	1.7
4	11.2	10.6	6.7	6.0
5	32.9	27.7	18.9	18.0
6	100.0	100.0	48.1	47.5
7			100.0	100.0

Comparisons were then made between the patterns of scores on the link items in 1975 and 1980. Table 7.7 shows the distributions of students for the sub-tests comprising the six link items for 10-year-old students and the seven link items for 14-year-old students. The distributions are in the form of cumulative percentages of students responding correctly to the items. The performance level required of 10-year-old students was five items correct. This level of performance was reached by 89 per cent of the 10-year-old students in 1975 and by 89 per cent again in 1980. It is clear that the proportions of 10-year-old students across the distribution were very consistent, a further indication that performance on the link items was similar in both test administrations. The performance level of five items correct out of seven was reached by 93 per cent of 14-year-old students in 1975 and by 94 per cent in 1980. Consistency in the proportions of 14-year-old students across the distribution was again marked, indicating how closely performance on the link items in 1980 matched that of 1975. It was concluded that there was no significant difference between performance on the link items in 1975 and 1980, either for 10-year-old or 14-year-old students.

Another important area of comparison between the 1980 and 1975 test administrations was performance on the items associated with a simulated newspaper. The format of the newspaper was identical on both occasions and the items were similar or common. As has been reported, the 1980 performance of 10-year-old students on six newspaper items, three of which were common items, matched their performance on the total test (71 per cent were successful in each case). This result contrasts with the poor performance of 10-year-old students in 1975: 27 per cent only reached the required level on the newspaper sub-test, a result which significantly depressed their performance on the total test (53 per cent attained mastery here). It was observed as a result of the 1975 study that 14-year-old students had performed very satisfactorily on tasks associated with reading a newspaper. The 1980 data indicated that an even higher level of performance had been reached on similar tasks. This conclusion was based on performance in the 13-item sub-test Usage (Community), which contained four common items. In this sub-test, 75 per cent of 14-year-old students reached the required level of 10 items correct, compared with 64 per cent of students who were successful (11 items correct out of 14) in the equivalent sub-test in 1975.

There was a further though indirect, means of comparison at each age level between the 1980 ASSP reading test results and those of the 1975 ASSP program. The comparison was achieved by relating performance on the comprehension of a prose passage to the field-testing exercise, in which an attempt was made to establish comparative difficulty levels for the prose passages used in 1975 (the *Tailor Bird* and *Paracutin* passages) and new passages, also taken from the IEA Reading Project of 1970. To this end, every student in the field-testing exercise was assessed on the appropriate 1975 prose passage as well as a new one. In the case of the 10-year-old students, a new passage, the *Sundew*, was chosen for the final test. A comparison of the 1980 results with those of 1975, after adjusting for the relative difficulties of the two passages obtained during field-testing, indicated that the 10-year-old students in 1980 appeared to have reached the same level of performance on the comprehension of a prose passage as their coevals had done in 1975. In the case of the 14-year-old students, a new passage, the *Paper* passage, was chosen for the final test. When the 1980 results on the new passage were compared with those for 1975, after again adjusting for the relative difficulties of the two passages obtained during field-testing, it would appear that the 14-year-old students in 1980 performed better on the reading of a prose passage than their coevals had done in 1975. This superiority was of the order of five per cent more students correct per item.

Reading: Differences Between Sub-Groups

Sex Differences in Performance

At the 10-year-old level, a higher proportion of female students gave the correct response to every item. There was an average of five per cent difference between the sexes on each item, with the greatest difference (nine per cent) resulting from Item *10*, which was a complex sentence containing a hypothetical condition relating to a girl and her friend.

Female students also performed better than male students on most items in the test for 14-year-olds. Results on six items were equal, and the male students were superior in only five items, by an average of three per cent. Of those five items three concerned the comprehension of a prose passage; Item 16, about the main idea of a paragraph, gave male students an advantage of eight per cent, although overall performance on this item was low. Female students performed better on the other 24 items by an average of three per cent on each item; their best performance came on Items 25 and 34, assessing reference skills in the newspaper, and here they were superior to male students by 10 per cent and 11 per cent respectively.

These findings are similar to those of the 1975 reading tests. Details of sex differences in performance in 1975 may be consulted in Keeves and Bourke (1976:98).

Original Students/Replacement Students - Differences in Performance

When students named for the tests at a given school were unable to take part in the test program, their places were taken by one or more of the five reserve students. Seven per cent of the achieved sample of 10-year-old students and 11 per cent of the 14-year-old students were replacements; these numbers were not considered significant in comparing performance between the two sub-groups. Nevertheless a comparison was made of the proportions of original students and replacement students who attained mastery on the total reading test scores. At the 10-year-old level, the figures were 71 per cent for original students and 66 per cent for replacement students; this difference was not significant. The figures for 14-year-olds were 83 per cent in each case.

Original School/Replacement School - Differences in Performance

Certain schools in the sample declined to take part in testing and were replaced early in the program. Other schools withdrew later and were replaced if time allowed. Investigations revealed that 70 per cent of 10-year-old students from original schools gained mastery, compared with 78 per cent of students from replacement schools. However, when the 71 per cent performance level of the total sample was taken into account, the difference was not considered significant. Results for 14-year-old students were only one per cent different, the figures being 83 per cent and 84 per cent respectively.

Location Differences in Performance

The results of students attaining the mastery level on the total scores in the reading tests were compared with the location of their schools. At neither age level was there any significant difference in the performance of metropolitan students compared with that of non-metropolitan students. Actual figures were 72 per cent and 69 per cent respectively for the 10-year-old sample and 84 per cent and 82 per cent respectively for the 14-year-old sample.

Summary

Two samples of Australian students aged respectively 10 and 14 years took reading tests which emphasized the comprehension of printed material of a kind commonly found in school classrooms. For the total test, a mastery level of 22 items correct out of 27 was required of the 10-year-old students and 71 per cent of students reached that level of performance. The corresponding mastery level for 14-year-old students was set at 28 items correct out of 35 and 83 per cent of the older students reached that level.

The pattern of performance in both age groups was similar in several respects. At both age levels high proportions of students showed their ability to read simple and complex sentences with understanding: the figures for the 14-year-old students were 100 per cent and 99 per cent respectively and, for the 10-year-old students, 99 per cent and 81 per cent. Lower proportions of students at both age levels (83 per cent of 14-year-old students and 70 per cent of 10-year-old students) mastered the reading of continuous prose written in an expository style. High proportions of the 14-year-old students (97 per cent) were successful in using their reading abilities to gain information from reference materials, especially in the classroom situation. There was evidence that the reference skills of the younger students were not so well developed, as fewer than three quarters were successful in the relevant sub-test. Their lower performance may indicate the greater dependence of 10-year-old students on teachers for the acquisition of knowledge. A newspaper was used as stimulus material for one section of the tests, as it was in 1975, and, on this sub-test of six straightforward items, 10-year-old students performed at a level which matched their performance on the test as a whole. Some three quarters of the 14-year-old students were successful on the comprehension of newspaper articles and a much higher proportion of these students showed they could use reference skills relating to a newspaper.

As was expected, 14-year-old students performed better than the 10-year-old students on three common sub-tests. On anchor items assessing performance in fundamental skills such as knowledge of the alphabet and reading simple sentences, differences between the age groups were minimal. Greater differences between the age groups appeared in performance on more difficult items, involving complex sentences, reference skills and newspaper reading, in which tasks the wider experience of 14-year-old students contributed to their higher performance. These differences ranged from four per cent on newspaper reading to 28 per cent on reference skills.

A comparison of results in the 1975 and 1980 test administrations was of great interest. In this respect, the mean results for the link items did not reveal significant differences in performance at either age level. However, indirect links with the 1975 ASSP programs had been made through procedures adopted during field-testing. It appeared that 10-year-old students had performed at the same level as their age mates in 1975 on the comprehension of a prose passage, while 14-year-old students had bettered the performance of their age mates at that time, approximately five per cent more students being correct on each item.

It was also possible to make a cross-national comparison of performance on the comprehension of the prose passage used in the 14-year-old reading test. This was done by means of data from the 1970 IEA study of Reading Comprehension for several other English-speaking countries. These data showed that the 14-year-old Australian students in 1980 had performed significantly better than students in England on the passage in question. However their performance was not significantly different from that of students in the USA, on the one hand, or from that of New Zealand students, on the other. It was noted as a result of the 1975 ASSP study that 14-year-old New Zealand students in 1970 were superior in reading comprehension to other students in the English-speaking world. It would be of interest to discover whether present-day 14-year-old New Zealand students have maintained the superiority in reading comprehension that was shown by their age mates a decade ago.

An interesting comparison was made between performance in 1975 and 1980 on items associated with a simulated newspaper, the format of which was identical in both test administrations. It is of note that the performance of 10-year-old students in this sub-test in 1980 equalled their performance on the total test. By contrast, the poor performance of 10-year-old students on the newspaper sub-test in 1975 depressed their performance on the total test. The 1980 results also indicated that 14-year-old students had performed better on tasks associated with reading a newspaper than their coevals in 1975, at which time it was considered that 14-year-old students had performed very satisfactorily in this area.

Comparisons have been drawn between performance on similar sub-tests in 1980 and 1975. A final point of comparison concerns the percentages of students gaining the required level on the total test score. Cutting scores on both occasions were equivalent and it is interesting to observe that 18 per cent more 10-year-old students and 11 per cent more 14-year-old students gained mastery on the total test score in 1980 than in 1975. It is clear that the improved performance on newspaper reading, especially by the 10-year-old students, made a substantial contribution to these higher levels of performance on the total tests.

Investigations into the performance of selected sub-groups in the student sample disclosed no significant differences with respect to the status of students (original or replacement), the status of schools (original or replacement) and location of schools (metropolitan and non-metropolitan). However there were significant differences in results according to sex. Female students out-performed males at both age levels: 10-year-old female students were better than male students by an average of five per cent on every test item and 14-year-old female students were better on 24 of the 35 items, by an average of 3 per cent.

The group-testing approach and constraints of time limited the scope of the 1980 ASSP reading assessment. However there is evidence from the study of the performance of Australian students aged 10 and 14 years in three important skill areas related to their development as independent learners, namely comprehending continuous prose, using reference skills and reading adult-oriented publications such as a newspaper. It is clear that many students in both samples were successful in the three areas and their overall attainment in reading may well be very high indeed. It is equally clear that other students did not perform well in these areas; in practical terms, such students would be likely to find difficulty in reading school text-books, seeking information on their own account and coping with the print media.

CHAPTER EIGHT

PERFORMANCE ON THE WRITING TESTS

The view of writing upon which the development of the Writing Tests was based is that of a purposeful activity in which several systems of knowledge or skills are integrated in response to writing demands. Two main areas from which writing demands arise were delineated. The first pertains to individual needs in the development of ideas, the structuring of experience and the expression of personal feelings and the second to demands society might make at a personal, group or institutional level.

Attempts were made in the construction of the tests to ensure that tasks represented as far as possible a range and variety of writing activity in both these areas. In addition tasks were selected for appropriateness to age level so that some tasks were set for only one age level and some for both. However it was recognized that in a test of approximately thirty minutes relatively few tasks could be set. Although some simple tasks require only a few minutes to complete, tasks in continuous prose such as writing opinions require much longer. So that a wider range of objectives could be tested two writing tests covering different tasks were given to approximately half the sample at each level. The development of the tests is explained in more detail in Chapter Two.

The results of student performance on writing tasks which are reported in this chapter are based on the weighted numbers for the total Australian sample. Types of scoring differed according to the nature of the task. For some tasks a number of sub-tasks were specified. From these sub-tasks criteria were selected which determined what were the necessary parts of a correct response to a particular task. Sub-tasks were scored on a correct/incorrect basis. For a student to successfully complete a task correct responses had to be made to a number of specified sub-tasks. In these tasks, which are reported individually, criteria for a correct response are outlined. For each task the percentage of correct responses for the population tested has been given for the completed task and for sub-tasks. Comparisons of performance between tasks and age levels and with the 1975 study have been made where appropriate.

For other tasks, for instance those requiring answers in continuous prose, global scoring was considered appropriate. Two methods of global scoring were used; marking by impression based upon general criteria and marking by primary trait. These methods have been described in detail in Chapter Two. In this chapter tasks which were globally scored are reported as a group.

Finally, performance on a selection of tasks was examined according to the background variables of sex, school location, year level, and original and replacement schools and students.

Student Performance on Criterion-Referenced Tasks

Writing a Telephone Message

This task was assessed by the item *Take a message* in tests 10W/B and 14W/A. Students were asked to write down a message for a family member based on a spoken communication. The teacher read out to the students the following statement, which purported to have been heard over the telephone:

> This is John Spencer calling. I'm in an awful hurry. Something's come up. I want to leave a message for Betty. My name is John Spencer, she'll know who I am, (Spell the next word) S P E N C E R. (Pause) I have to go to hospital for some tests tonight. (Pause) Tell Betty I shall meet her at the Post Office at 9 o'clock. Got that? The Post Office, 9 o'clock tonight. Thank you. Goodbye.

From a number of possible elementary sub-tasks both of the following were considered necessary for the satisfactory completion of the task.

1 Giving the name of the person who left the message.
2 Giving the time and place of the meeting.

Results for the performance of students on this task are shown in Table 8.1.

As might be expected 14-year-old students performed better than 10-year-olds on the whole task and on each sub-task. Whereas 92 per cent of 14-year-old students completed the whole task successfully only 69 per cent of 10-year-old students were able to do so. Differences between the age levels were comparatively small on the sub-task which required the name of the caller to be given. Of the 10-year-old students 94 per cent completed this sub-task successfully and 99 per cent of 14-year-olds. Differences in performance at the two age levels were greater in the sub-task which required the time and place of the meeting to be given where 92 per cent of 14-year-olds and 70 per cent of 10-year-old students were successful.

This task was also administered to the two age groups in 1975. In that year 73 per cent of 10-year-olds and 89 per cent of 14-year-olds completed the task successfully. The differences in performance at each age level in

Table 8.1 Comparison of Performance in Writing a Telephone Message by Students Aged 10 and 14 Years in Australian Schools in 1975 and 1980

Sub-tasks	% Correct 10-Year-Olds	% Correct 14-Year-Olds
1980		
1. Name of caller	94	99
2. Time and place of meeting	70	92
Completed task (1 and 2)	69	92
1975		
Completed task	73	89

1975 and 1980 were not statistically significant according to the required differences shown in Table 3.6. However in both 1975 and 1980, 10-year-old students performed considerably less well than 14-year-old students, and this difference was particularly noticeable in the details of the message. Possible explanations for the large difference include the greater familiarity of 14-year-old students with the type of task, and a better facility in integrating listening with writing skills.

Writing a Personal Message

This task was assessed by the item *Write a message* in test 10W/A, and by a different item of the same name, in test 14W/B. From information presented in Figure 8.1 below 10-year-old students were asked to leave a message for their mother telling her where they had gone and when they would be home.

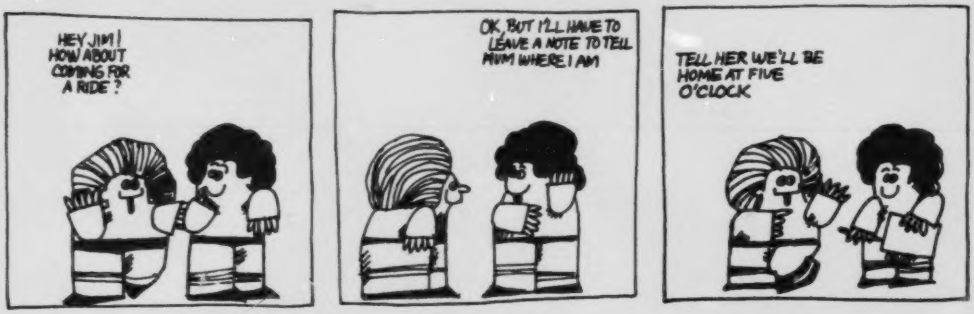

HEY JIM! HOW ABOUT COMING FOR A RIDE?
OK, BUT I'LL HAVE TO LEAVE A NOTE TO TELL MUM WHERE I AM
TELL HER WE'LL BE HOME AT FIVE O'CLOCK

Figure 8.1 Stimulus Material for Writing a Personal Message, 10-Year-Old Students

Table 8.2 Performance in Writing a Personal Message by Students Aged 10 Years in 1975 and 1980 and Students Aged 14 Years in 1980, in Australian Schools

Sub-tasks	% Correct 10-Year-Olds		% Correct 14-Year-Olds
	1975	1980	1980
1. Mention where going	96	97	99
2. Give time of return	88	94	97
Completed task (1 and 2)	86	93	97

The following sub-tasks were considered necessary for the successful completion of the task.

1. Mentioning the ride, or the place to which they were going
2. Giving the time of return

The task which 14-year-old students were asked to complete was similar but the information and cartoon pictures were more appropriate for the age level. Students were asked to leave a message for their mother saying where they had gone and when they would be home. The criteria for successful completion of the task were:

1 Mentioning the place to which they were going.
2 Giving the time of return.

Table 8.2 shows that 93 per cent of 10-year-olds and 97 per cent of 14-year-olds completed the task successfully. Ninety-seven per cent of 10-year-olds and 99 per cent of 14-year-olds mentioned where they were going. Ninety-four per cent of 10-year-olds and 97 per cent of 14-year-olds gave the time of return. The students in the older age group performed better than those in the younger group but both did well. It would seem that this task, which required writing a personal message, has been mastered by almost all students at both age levels.

In 1975 10-year-old students attempted the same personal message as 10-year-old students in 1980 (see Table 8.2). Of students in 1975, 86 per cent completed the task successfully compared with 93 per cent in 1980. The difference of 7 per cent indicates a significant improvement according to the statistical requirements outlined in Chapter Three (see Table 3.6). Most of this difference is accounted for by the better performance in 1980 on the sub-task requiring the time of return to be given, which shows an improvement of 6 per cent.

The percentage figures in Table 8.2 show that performance of 10-year-olds in writing the personal message has improved between 1975 and 1980.

Writing Directions

The objective of writing instructions includes several important tasks, of which writing directions is one. Owing to constraints of time, it was decided that writing directions should be the only task tested in this area and that it should be confined to the 14-year-old students. Consequently the item *Write Directions* was included in test 14W/B. A simple street map was given to students who were then asked to write directions to enable a person to drive from the school grounds to a library shown on the map. For a student's response to be considered correct sufficiently precise directions had to be given to enable a person to reach the library from the school grounds. A minimum of four separate instructions was considered necessary for the most obvious route. Alternative routes were considered acceptable if the general conditions for a correct response, explained above, were satisfied.

Of students who attempted the task only 54 per cent completed it successfully. Because of the nature of the task there is no room for error in giving the instructions. In practical terms if one direction is inaccurate then accuracy in the rest will not enable a person to reach the destination required. However it is interesting to note that 80 per cent of students were able to give at least three accurate directions out of four. On the first instruction 29 per cent of students failed to indicate a right hand turn from the school grounds into the street and on the second instruction 27 per cent of students failed to indicate a necessary left hand turn. The two other instructions were completed satisfactorily by approximately 90 per cent of students.

In 1975 when the same item was used, 52 per cent of 14-year-old students completed this task. The consistently low performance on the two occasions indicates that this task, which involves writing four directions, presents difficulties for many students at this age level.

Writing the Main Idea and Giving Supporting Detail

These tasks were assessed at the 10-year-old level only by the item *Look at the picture story below* in test 10W/A. Students were asked to look at the picture story, shown in Figure 8.2, then provide a title and write what the man is saying in the fourth picture.

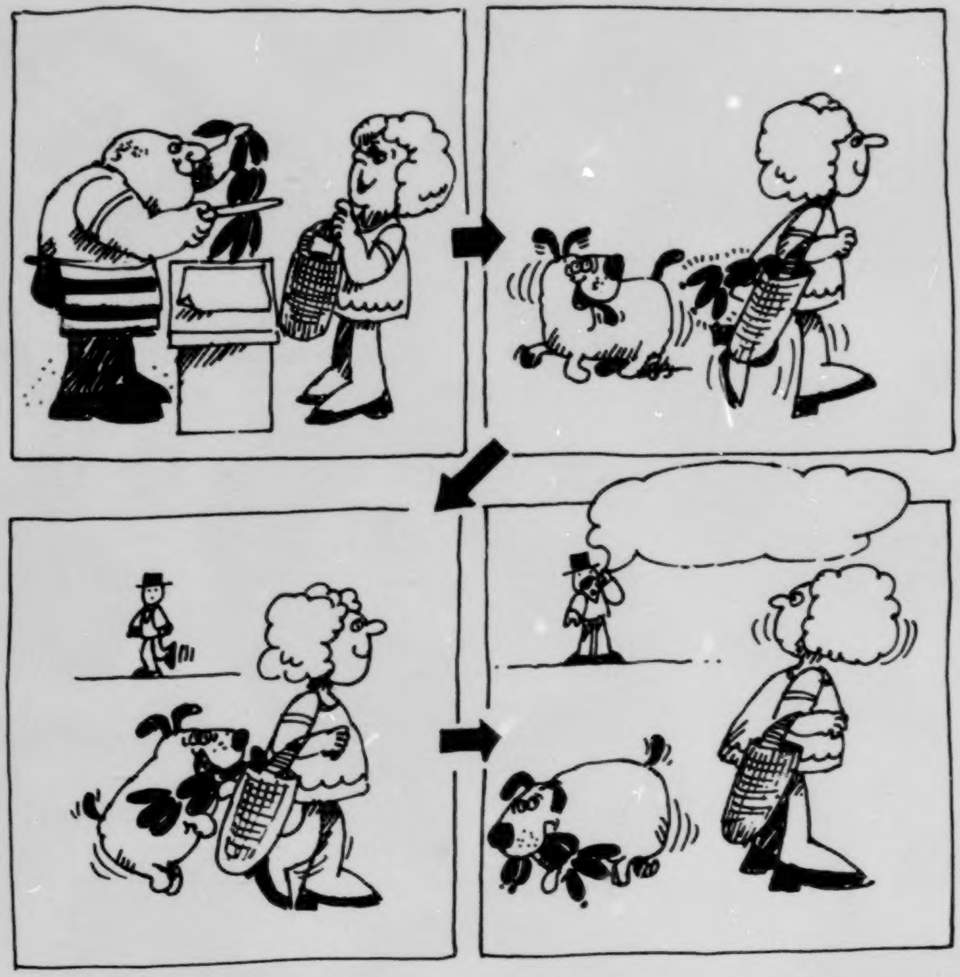


Figure 8.2 Stimulus Material for Providing a Comment and a Title, 10-Year-Old Students

To be acceptable the title had to summarize the point of the cartoon strip or refer to one of the major elements. The comment had to contain a phrase appropriate to the events in the picture. Most students were able to complete the separate tasks satisfactorily. The title was completed successfully by 83 per cent, the comment by 91 per cent of students.

This task was also given to 10-year-old students in 1975. A satisfactory title was given by 93 per cent of students and an appropriate phrase or comment by 86 per cent. It should be noted that 7 per cent of students omitted to fill in a title in 1980 and 7 per cent of students omitted to fill in a comment in 1975. These omissions suggest that students focussed on one task only and rather than being unable to complete the other task did not take it into account.

Writing Letters and Addressing Envelopes

Two kinds of exercise which involved writing letters and addressing envelopes were included in the Writing Tests. In one, which was administered to both age levels, students were required to write for information about a particular subject and to address an envelope. In addition, 14-year-old students were asked to write a letter to an organization offering to help in a youth program. They were also asked to address an envelope.

1. Writing a Letter for Information and Addressing an Envelope.

This task was assessed by the item *Write a letter for information and fill in an envelope* in tests 10W/B and 14W/B. Information was presented to the students in the form of an advertisement which is shown in Figure 8.3. The information purported to be about an organization called 'Save Our Seals'. Students were asked to write a letter requesting information about seals and address an envelope to an official of the organization.

S.O.S. SAVE OUR SEALS S.O.S.

Do you know what happens to seal pups?
Do you want to help them?
Write to Jim Fisher at S.O.S. for information
about our seals. Don't forget to tell us
your name and address.

Send your letter to:

Jim Fisher,
Save Our Seals,
86 Shell Road,
Rock Point, 5871.

Figure 8.3 Stimulus Material for Writing a Letter Requesting Information, 10- and 14-Year-Old Students.

Table 8.3 Performance in Writing Letters and Addressing Envelopes of 10- and 14-Year-Old Students in 1980

Task: Letter Requesting Information		
Sub-tasks	% Correct 10-Year-Olds	% Correct 14-Year-Olds
On the envelope		
A. Number & Name of Street	93	99
B. Town or Post Code	93	99
Envelope Essentials (A + B)	93	99
In the Letter		
C. Including return address	69	91
D. Including greeting	90	99
E. Information requested	85	96
F. Signature included	80	97
Letter Essentials (C + E + F for 10-Year-Olds	58	
(C + D + E + F for 14-Year-olds)		87
Completed Task (Envelope + Letter Essentials)	57	87

Task: Letter Offering Help	
Sub-tasks	% Correct 14-Year-Olds
On the envelope	
A. Title of Council Youth Organizer	97
B. Post Office Box Number	98
C. Town	97
D. Postcode	97
Envelope Essentials (B + C or D)	97
In the Letter	
E. Including return address	92
F. Including a greeting	92
G. Including an offer of help	97
H. Stating dates available	81
I. Mentioning interests	67
J. Including signature	95
K. Including telephone number	74
Letter Essentials (E or K + G + J)	89
Letter (E + F + G + J)	83
Completed Task (Envelope + Letter Essentials)	87

Several sub-tasks were identified for the letter and the envelope. From these sub-tasks criteria were selected in assessing the successful completion of the task. The choice of criteria was based on what the student would be expected to include to receive the information required. The following criteria had to be met for a response to be correct for 10-year-old students.

On the envelope 1. The number and name of the street (86 Shell Road)
2. The town or post code (Rock Point/5871)
In the letter 3. A return address
4. A request for information about seals or seal pups
5. A signature.

In addition, 14-year-old students were expected to include a greeting.

Table 8.3 shows the performance of students at both age levels. As well as showing the percentage of students who completed each sub-task and the whole task successfully it also includes separately the percentages of students who were correct in addressing the envelope and in completing the essentials of the letter. Students at both age levels performed well on that part of the task which required the envelope to be addressed. Of the 14-year-old students 99 per cent were successful as were 93 per cent of 10-year-old students. Differences in performance were greater for the satisfactory completion of the letter. Only 58 per cent of 10-year-old students gave the letter essentials compared with 87 per cent of 14-year-old students. At both levels the sub-task which was most frequently omitted was the inclusion of the return address in spite of the fact that this was specifically requested in the advertisement shown in Figure 8.3. From the results of this part of the test it appears that a high proportion of 14-year-olds were able to fulfil basic requirements in writing this letter.

In 1975 students at both age levels were required to write letters asking for information (see Table 8.4). These letters were different at each age level and were not repeated in 1980 so that direct comparisons are not possible across the whole task. However it is interesting to note that considerably more 10-year-olds used a greeting in 1980 (90 per cent) than in 1975 (63 per cent) and a larger proportion signed their names in 1980 (80 per cent) than in 1975 (71 per cent). It is possible that responses were influenced by the less formal approach taken in the stimulus materials in 1980 which made the sub-tasks more accessible. Corroborating this explanation is the fact that in the personal letter given to 10-year-old students in 1975, a greeting was used by 95 per cent and 79 per cent signed their name. It does

Table 8.4 Performance in Writing Letters by Students Aged 10 and 14 Years in 1975[a].

	10-Year-Old Students		14-Year-Old Students	
Selected Sub-tasks	Personal Letter % Correct	Letter requesting information % Correct	Letter requesting information % Correct	Letter requesting employment % Correct
A. Return address included	30	60	80	68
B. Greeting included	95	63	97	89
C. Purpose given	94	92	95	77
D. Signature included	79	71	95	91
Completed task	(C+D) 76	(A+C+D) 50	(A+B+C+D) 74	(A+B+C+D) 50

[a] Adapted from Tables 7.4, 7.5 and 7.6 in Keeves and Bourke (1976:76,78,79).

appear that the performance of 10-year-old students is affected adversely by formality and distance and that they perform better in letter writing skills when approaches to a task are personalized.

When comments were made on the trial tests the letter-writing exercise which 10-year-old students were asked to complete in 1975 was criticized by some for being dull and lacking interest whereas the exercise included in the 1980 writing tests was preferred. However in 1980 fewer students at this level specifically requested information (85 per cent) than in 1975 (92 per cent). The high level of interest in conservation, including the protection of seal pups, might have deflected the attention of a proportion of students away from the set task.

Of the sub-tasks attempted by 14-year-old students a return address was included by 91 per cent in 1980 and 80 per cent in 1975. A likely explanation for some of this difference is that whereas in the 1975 exercise a return address was not mentioned, in the 1980 exercise one was asked for in the stimulus materials. Performance in 1975 and 1980 on other sub-tasks, namely the inclusion of a greeting, a request for information and a signature, showed small variations and a high level of success. This was not the case at the 10-year-old level where there were considerable differences between sub-tasks, and differences in the order of levels of performance.

2. Writing a Letter Offering to Help and Addressing an Envelope

The stimulus material in Figure 8.4 was presented to 14-year-old students in the item *Write a letter and address an envelope* (Test 14W/A). Students were asked to write a letter offering to help in a Youth Program, give the dates when they would be available and information about themselves which would be useful to the recipient of the letter. In addition they were asked to address an envelope to the Youth Organizer. This advertisement, as can be seen, is more formal than the one for the letter requesting information. Also the address of the organization is not so obviously laid out.

From a number of sub-tasks the following were considered to be necessary for the successful completion of the task.

On the envelope

1 The Post Office Box Number.

2 The Town or Post Code.

On the letter

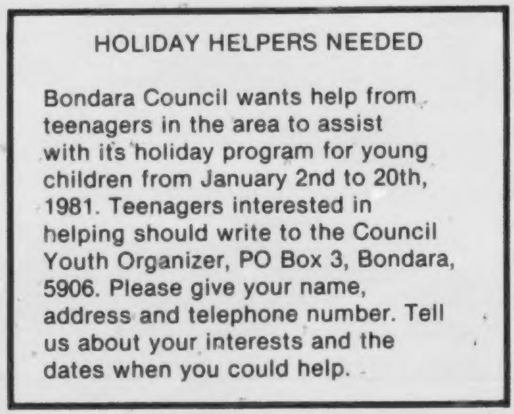
3 Inclusion of a return address or a telephone number.

4 Inclusion of an offer to help.

5 Inclusion of a signature.

Results are reported in Table 8.3.

Of the 14-year-old students who were given this task 87 per cent completed it successfully. Most performed very well on addressing the envelope, although the level of performance in this part of the task was slightly lower than in the equivalent part of the letter writing task requesting information. The more difficult format of the address presented in the Bondara advertisement probably contributed to this difference.



HOLIDAY HELPERS NEEDED

Bondara Council wants help from teenagers in the area to assist with it's holiday program for young children from January 2nd to 20th, 1981. Teenagers interested in helping should write to the Council Youth Organizer, PO Box 3, Bondara, 5906. Please give your name, address and telephone number. Tell us about your interests and the dates when you could help.

Figure 8.4 Stimulus Material in the Letter Offering to Help, 14-Year-Old Students.

In the letter itself, on the sub-tasks which required the inclusion of a return address, the purpose of the letter and a signature, performance was similar to equivalent sub-tasks at the 14-year-old level in the letter requesting information. However in the letter offering help a greeting was more frequently omitted. In the stimulus materials the recipient of the letter was given a title (Council Youth Organizer), whereas in the letter requesting information a personal name was given (Jim Fisher). It is also worthwhile noting in passing that performance on this sub-task was slightly higher than that on a similar sub-task in the letter for employment in 1975, where again a title rather than name was given (see Table 8.4). It appears that the formality in the use of a greeting could have presented a problem to some students at this age level.

Other information required in the letter was given by a smaller percentage of students: 81 per cent of students gave the dates on which they were available, and 67 per cent included a mention of their interests. Both these sub-tasks were considered of lesser importance in establishing basic requirements for the letter although their omission indicated to some extent a lack of awareness of some students about the audience for whom the letter was purported to be written and about the relevance of supplying supporting information about themselves.

Performance on a combination of the four sub-tasks; include a greeting, a signature, a return address and the purposes of the letter, was slightly but not significantly lower (83 per cent) than the equivalent combination (87 per cent) in the letter requesting information.

Writing a letter offering help was also scored globally, by the primary trait method which involved experienced teacher-markers assigning a grade according to an important focus of the task (see Chapter Two). In this case social cognition was considered to be of focal importance. The marker assigned a score according to the degree of awareness shown by the student in the selection and presentation of information to a known but formal audience in a manner which shows cognizance of the conventions of letter writing. The global score gave an indication of quality of response, whilst the letter essentials score showed performance on specific criteria. A correlation of 0.36 between the two scores shows that only approximately 13 per cent of the variance of the global score could be accounted for by performance on the letter essentials. Clearly whilst basic requirements formed a small part in qualitative assessment other important aspects of the writing system were also being considered by the markers, although in a less specific way.

Writing a Report

One of the elements involved in writing a report was considered to be the ability to describe accurately a series of events. Students at the 10-year-old level were given six simple drawings depicting the growth of beans, from a person planting bean seeds to a family eating the beans. Students were asked to say what happened in the pictures. For a satisfactory response an appropriate reference had to be made to every one of the frames. Single sentences or phrases covering more than one of the frames were acceptable.

It was found that 81 per cent of students made a reference to all of the frames and 94 per cent to at least five of the six frames. When each frame was considered separately the range of students responding appropriately varied from 89 per cent to 98 per cent. Ninety-eight per cent of students reported the first frame and 97 per cent the last. The major omission occurred in frame four which depicted a person chopping beans. This frame was not referred to by 11 per cent of students.

This task was also scored globally. A correlation co-efficient of 0.31 shows that the total score arrived at by criterion-referenced marking was related in a minor way to the global assessment made. As was suggested in reporting the results on the letter offering help written by 14-year-olds, quality of written performance included more aspects of the writing system than the fulfilment of basic requirements.

Summary of Performance on Criterion-referenced Writing Tasks

At the 10-year-old level performance between tasks in which criteria were specified showed a considerable variation. From Table 8.5 it can be seen that successful completion of tasks varied from 57 per cent to 93 per cent, a difference of 36 per cent. The task most successfully performed was writing a personal message to a family member (93 per cent). Many students found difficulty in successfully completing the letter for information, to which 57 per cent responded correctly, and taking a telephone message, to which 69 per cent responded correctly. At the 14-year-old level successful responses to completed tasks varied from 54 per cent to 97 per cent, a difference of 43 per cent. As with the 10-year-old students, writing a personal message to a family member was very well done. The most difficult task for 14-year-old students was writing directions. Performance on this task showed a far lower success rate than for any other item.

Table 8.5 Performance on Completed Criterion-Referenced Tasks by Students Aged 10 and 14 Years in 1980

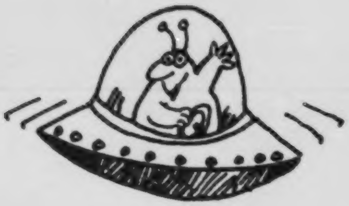
Task	% Correct	
	Age 10	Age 14
Writing directions	not tested	54
Writing a message	93	97
Taking a telephone message	69	92
Writing a report	81	not tested
Writing a letter for information and filling in an envelope	57	87
Writing a letter offering to help and addressing an envelope	not tested	87
Writing the main idea and giving supporting detail	79	not tested

In the two tasks which were given to both age groups, the 14-year-old students showed greater facility. Of the 14-year-old students 92 per cent successfully completed the telephone message, compared with 69 per cent of 10-year-old students. Writing a letter requesting information and addressing an envelope was completed successfully by 87 per cent of 14-year-old students, compared with 57 per cent of 10-year-old students.

When performance in 1975 and 1980 was compared, taking a telephone message and writing directions showed a slight, but not significant increase in the percentage of correct responses at the 14-year-old level in 1980. For 10-year-olds in 1980 there was a slightly lower but not significantly different level of performance in taking a telephone message and a significant improvement in writing a personal message. For the task of writing a letter for information, the common sub-tasks of writing a greeting and giving a name were completed successfully by higher proportions of the 10-year-old students in 1980. Performance did not differ for the 14-year-olds between 1975 and 1980.

Student Performance on Globally Scored Tasks

Three tasks at each age level were scored globally. Two methods of global scoring were used. One involved marking by impression. This method of marking, which did not involve the allocation of a mark to a specific criterion, was used at the 10-year-old level for the tasks *Write a Report*, *Write About Yourself* and *Write a Story*, and at the 14-year-old level for the tasks *Write a Story* and *Write Your Opinion*. The other global method, primary trait marking, which focusses on a student's ability to gain a desired effect



This is Riki.
He is your friend.
BUT
He plays tricks
and then disappears.

Figure 8.5 Part of the Stimulus Material for the Story, 10-Year-Old Students.

with an audience, was used at the 14-year-old level for the task *Write a Letter Offering Help*. These methods are placed in context in Chapter Two. Reliability estimates for the global scores allocated to each task have been reported in Appendix Five.

In addition to the global method of scoring for these tasks, quantitative analyses of the number of words written and the percentage spelling error per hundred words were undertaken. Examples of scripts with an indication of the score obtained can be found in Appendix Four.

Description of Tasks at the 10-Year-Old Level

Write About Yourself. This task was designed to give students the opportunity to write about their experiences. A number of photographs showing young people engaged in various activities were given. The students were asked to look at these and then write about the things they enjoyed doing.

Write a Story. The stimulus material for this task was based on two drawings which attempted to place an imaginary creature in a situation familiar to students. The creature was introduced as shown in Figure 8.5.
The second drawing showed the creature with a family at home. Students were then asked to write a story about what happened when it visited their home.

Write a Report. For this task, students were asked to describe a sequence of drawings depicting the growing and preparation of beans. The report was also marked by reference to specific criteria, and the assessment of performance according to these criteria has been described earlier in this chapter.

Description of Tasks at the 14-Year-Old Level

Write a Story. At the 14-year-old level students were asked to write a story about three friends. The task was introduced as shown in Figure 8.6.

The degree to which students made use of the information given was a matter for their discretion. Marking was based on the quality of the response, not on the inclusion or exclusion of the stimulus material.

Write Your Opinion. For this task the stimulus materials were based on a fictitious newspaper report of a speech which criticized lack of parental control over the younger generation. Students were then asked to write their opinion about teenage friends.

Write a Letter Offering to Help. For this task, students were asked to write a letter offering to help in a youth program. Student performance on writing this letter, marked according to specific criteria, has been reported earlier in this chapter.

Scores, Number of Words and Spelling Errors

Scores. For all these tasks two independent markers awarded a mark from 1 to 5 on the basis of general criteria outlined in Chapter Two. These scores were then totalled. The score referred to in this chapter is the total score obtained from the two marks which ranged from 2 to 10.

The similarity in the means and standard deviations of the scores, reported in Table 8.6, indicates that for all tasks there was a similar distribution of scores over the range. At the 10-year-old level the task *Write About Yourself* had a slightly lower mean. At the 14-year-old level for *Write Your Opinion* the mean score was approximately 0.5 lower than for the two other tasks.

Number of Words. The number of words written by students in responding to each of the globally-scored tasks varied considerably as can be observed in Table 8.7. For convenience of reporting, the number of words written were grouped into five categories which were designed to approximate a normal

In the past week too many pets have disappeared from the area. Twelve racing pigeons, two ponies and a talking cockatoo are among the missing animals.

Three friends decide to find out what is going on. They are very suspicious of the white truck parked outside the empty shop.

■ **Write a story about the three friends. (You can be one of the characters if you wish.)**

Figure 8.6 Stimulus Material for the Story, 14-Year-Old Students.

Table 8.6 Means and Standard Deviations of Global Scores, Number of Words and Spelling Errors per 100 Words of Writing Tasks of 10- and 14-Year-Old Students

Task	Impression Score		Number of Words		Spelling Errors	
	Mean	SD	Mean	SD	Mean	SD
10-Year-Old Students						
Write a Report	6.1	1.9	50	19	2.8	3.8
Write About Yourself	5.9	1.9	74	47	4.5	5.6
Write a Story	6.1	2.0	110	46	4.4	4.6
14-Year-Old Students						
Write a Letter	5.9	2.1	70	24	1.6	1.7
Write Your Opinion	5.4	2.1	144	67	2.4	3.2
Write a Story	5.9	1.9	267	82	1.6	1.7

distribution. Consequently the lowest category (Category 1) included approximately 10 per cent of the students and Categories 2 to 5 included respectively 20 per cent, 40 per cent, 20 per cent and 10 per cent of the sub-samples of students. The number of words in each category for each of the six tasks is shown in Table 8.7.

It was expected that the nature of the task and the time allowed would produce differences in the number of words written for each of the three globally-scored tasks. It should be noted that time allocations were exactly the same for *Write a Story* and *Write About Yourself* for the 10-year-old students and approximately the same for *Write a Story* and *Write Your Opinion* for the 14-year-old students. At each age level the third of the globally-scored tasks, that is *Write a Report* for 10-year-old students and *Write a Letter* for 14-year-old students, received approximately half the time allocation of the other two tasks.

Investigations showed that, at the 10-year-old level, there was a considerable difference in the number of words produced in each category. For *Write a Report* and *Write About Yourself* the number of words produced in Categories 1, 2 and 3 was similar. However, for the task *Write About Yourself*, there was a wide range of words in each of categories 4 and 5. This range in the number of words produced in the higher categories was reflected in the standard deviation reported in Table 8.6. From the means shown in Table 8.6

Table 8.7 Categorization of Number of Words Written by 10- and 14-Year-Old Students on Globally Scored Writing Tasks

Number of Words	Categories of Words Written				
	1 (10%)	2 (20%)	3 (40%)	4 (20%)	5 (10%)
10-Year-Old Level					
Write a Report	0-30	31-40	41-55	56-75	76-174
Write About Yourself	0-25	26-45	46-90	91-135	136-702
Write a Story	0-50	51-80	81-135	136-170	171-284
14-Year-Old Level					
Write a Letter	0-40	41-55	56-80	81-100	101-185
Write Your Opinion	0-65	66-105	106-175	176-240	241-474
Write a Story	0-165	166-220	221-310	311-370	371-632

it can be seen that the 10-year-old students tended to write more for *Write a Story* than for either of the other two items. It is also clear that students wrote the least number of words for *Write a Report*.

At the 14-year-old level, students showed a capacity to write more words for each of the globally-scored items than 10-year-old students. As with the younger age group, 14-year-old students tended to write more in telling a story than for either of the other two tasks; it was also found that 14-year-old students wrote the least number of words for the task *Write a Letter*.

Spelling Error Rate. To reduce the workload involved, a random sample of only 10 per cent of the student responses on the globally scored tasks were scored for spelling errors. A spelling error rate was then obtained by counting the number of errors made for every hundred words written. This was done for both the 10- and the 14-year-old students.

Of the three tasks at the 10-year-old level for which results are shown in Table 8.6, the students had a substantially lower mean spelling error rate in the report than for either of the other two tasks. For all tasks the standard deviation was high. Table 8.8 shows the distribution of the spelling error rate in six categories. From this table it can be seen that 43 per cent of students made no spelling error in the report compared with 23 per cent for the task which asked them to write about their experiences and 13 per cent for the story. Although between 65 and 79 per cent of the 10-year-old students made four or fewer spelling errors per 100 words, the remainder were spread over a greater range.

Table 8.8 Spelling Error Rate of 10- and 14-Year-Old Students on Globally Scored Writing Tasks

Task	Number of Spelling Errors					
	0 %	1-2 %	3-4 %	5-6 %	7-10 %	10+ %
10-Year-Old Level						
Write a Report	43	16	19	9	8	5
Write About Yourself	23	21	21	12	13	10
Write a Story	13	31	20	13	11	12
14-Year-Old Level						
Write a Letter	25	54	15	5	1	0
Write Your Opinion	23	42	20	7	5	2
Write a Story	25	54	15	5	1	0

At the 14-year-old level the range of spelling errors was less and a higher percentage of students made fewer errors per hundred words than 10-year-old students. For the task which required students to write their opinion, the mean spelling error rate was higher than for the two other tasks.

Fourteen-year-old students made fewer errors for every hundred words written than 10-year-old students, although there was some variation between tasks at both levels. However it was found that the correlation between the spelling error rates on two tasks at the 14-year-old level was almost perfect. For these two tasks, *Write a Letter Offering Help* and *Write a Story*, the nature of the tasks, which were dissimilar, did not affect the rate of spelling error.

Relationships Between Global Scores, Number of Words and Spelling Error Rate

The relationships between global scores, number of words and spelling error rate were investigated by means of correlation coefficients. Whilst the correlations between global scores and the number of words that students wrote, reported in Table 8.9, were all significant at the 5 per cent confidence level, the percentage of common variance differed from task to task. The correlation between global scores and number of words was more consistent at the 10-year-old level, where it ranged from 36 per cent to 42 per cent, than at the 14-year-old level. It is interesting to note that

Table 8.9 Correlations of Global Scores, Number of Words and Spelling Error Rate for 10- and 14-Year-Old Students

Correlations (r) Task	Global Score with Number of Words		Global Score With % Spelling Error		% Spelling Error with Number of Words	
	r	N	r	N	r	N
10-Year-Old Level						
Write a Report	0.64	2697	-0.37	552	-0.16	552
Write About Yourself	0.60	2695	-0.46	552	-0.18	558
Write a Story	0.65	2761	-0.57	565	-0.43	565
14-Year-Old Level						
Write a Letter	0.38	2523	-0.49	510	-0.28	510
Write Your Opinion	0.63	2524	-0.42	501	-0.33	501
Write a Story	0.47	2540	-0.54	510	-0.29	510

the correlations for the two tasks *Write a Story* (0.65) and *Write About Yourself* (0.60) were similar to those found in the 1975 study and reported as 0.62 and 0.58 respectively by Bourke and Keeves (1977:113).

At the 14-year-old level the percentage of common variance between the global score and number of words for the task of writing a letter was 14 per cent compared with 40 per cent for the task of writing opinions. Clearly the quality of the completed task was related less to the number of words in the letter, which had precise ends, a particular structure and a particular and formal audience, than in either the story, which allowed more freedom of choice within a number of conventions and was firmly based in concrete experience, or in writing opinions which was the most unstructured and abstract task.

The negative product moment correlations for global scores and spelling error rate reported in Table 8.9 were significant at the 5 per cent confidence level for each task. Students who made fewer errors tended to have a higher global score. The percentage of common variance between the global scores and spelling error rate was highest for the story at both age levels. It was lowest for the task, *Write a Report* at the 10-year-old level, which also had the lowest spelling error rate, at this age level, with a large percentage of students making no error.

At the 10-year-old level the correlation between the spelling error rate and number of words produced was significant at the 5 per cent confidence level for the story, which indicated that the more words students wrote the fewer errors per hundred words they made. However on the two other items there was no significant correlation between words produced and spelling error rate.

At the 10-year-old level for the report, the fewest words were produced and the fewest spelling errors were made for each hundred words written. This task was the most highly structured of the globally scored tasks. It followed a concrete sequence of events for which a basic and relatively limited vocabulary was sufficient. The tasks *Write About Yourself* and *Write a Story* imposed no structural framework upon the students. Both in 1975 and in 1980, the mean number of words written was higher for story writing than for writing about personal experience. Part of the explanation for this might be that students were more experienced in the former, so their word fluency was greater.

At the 14-year-old level students performed least well on writing their opinions, the most abstract and least structured task. Less was written for this task than for writing a story and it had the highest spelling error rate. The correlation between the global score and the number of words that a student wrote was highest for this task, at this age level.

Background Variables

In addition to reports on performance of the total samples of students on the writing tasks, further analyses of performance have been made according to background variables. Some of these have been selected because the comparisons were thought valuable for an understanding of the performance of students in Australian schools. Of this kind of variable the results for sex, school location and year level have been reported.

The two other background variables on which performance has been reported are those which would show the differences, if any, made to the reported performances on items by replacement schools and students. Original, replacement and total percentage performance have been given for both schools and students.

A selection of tasks has been reported for each variable and the selection has been mostly confined to tasks scored according to specific criteria.

Sex. In general for tasks scored according to specific criteria the differences in performance of male and female students were greater at the 10-year-old than at the 14-year-old level. Table 8.10 shows the percentage of successful students on some of these tasks. The tasks selected for presentation are those which were attempted at both age levels (the telephone message and the letter) and those which were very similar (the personal message).

Table 8.10 Distribution of Correct Responses on Three Writing Tasks by Male and Female Students Aged 10 and 14 Years

Students	Telephone Message		Personal Message		Write a Letter	
	% Correct	N	% Correct	N	% Correct	N
10-Year-Old						
Males	66	1528	90	1313	51	1313
Females	[illegible]	1211	95	1386	67	1216
14-Year-Old						
Males	91	1289	96	1308	82	1308
Females	94	1128	98	1234	93	1234

At the 10-year-old level it can be seen that females performed considerably better than males as was the case for all tasks of this type. Also it should be noted that the easier the task the smaller the difference. Whilst females tended to perform better at the 14-year-old level this was not so for all tasks. In the task, *Write Directions*, males performed slightly better (54 per cent) than females (52 per cent). Other differences at the 14-year-old level were slight except for the task which required students to write a letter requesting information, where there was an 11 per cent difference, 93 per cent of females responding correctly and 82 per cent of males.

Table 8.11 shows the percentage distribution of scores of male and female students on tasks which were marked globally. At both age levels a greater proportion of males received scores in the three lower categories for each task. Conversely a greater proportion of females received scores in the three higher categories. At the 14-year-old level the differences were greater in low scores between males and females than at the 10-year-old level. The greatest difference was in the task, *Write Your Opinion*.

In tasks which were marked according to specific criteria, differences in the proportions of correct responses were greater between male and female students at the 10-year-old level. In those tasks which were scored globally, however, the differences between the proportions of males and females receiving the three lower scores were greater than at the 10-year-old level. In both cases the results favoured the female students.

School Location. Students were classified according to the area in which they attended school, either in a capital city (metropolitan) or elsewhere (non-metropolitan). Results are reported in Table 8.12. Differences were

Table 8.11 Distribution of Global Scores by Sex at the 10- and 14-Year-Old Levels

Task	Global Score 2 %	 3 %	 4 %	 5 %	 6 %	 7 %	 8 %	 9 %	 10 %
10-Year-Old Level									
Write a Report:									
Males	5	6	16	22	23	14	6	5	3
Females	2	2	7	14	20	24	13	11	6
Write About Yourself:									
Males	6	8	17	18	21	15	7	6	2
Females	2	4	11	15	22	20	13	9	3
Write a Story:									
Males	6	7	17	17	20	16	7	6	3
Females	3	3	10	12	21	20	13	11	7
14-Year-Old Level									
Write a Letter:									
Males	11	11	13	18	17	14	9	5	1
Females	3	6	5	14	19	22	13	12	6
Write Your Opinion:									
Males	15	15	14	21	14	10	5	4	1
Females	3	7	10	15	21	19	10	9	5
Write a Story:									
Males	8	11	15	22	15	15	8	4	2
Females	2	4	6	17	18	22	14	13	15

Table 8.12 Distribution of Correct Responses of 10- and 14-Year-Old Students on Three Writing Tasks by School Location

Location	Telephone Message % Correct	 N	Personal Message % Correct	 N	Write a Letter % Correct	 N
10-Year-Old Students						
Metropolitan	66	1325	94	1558	60	1349
Non-metropolitan	73	1414	91	1142	57	1419
14-Year-Old Students						
Metropolitan	92	1402	97	1493	86	1493
Non-metropolitan	93	1016	96	1049	89	1049

Table 8.13 Distribution of Correct Responses of 10-Year-Old Students on Four Tasks by Year Level

Task	Telephone Message		Write a Letter		Personal Message		Write a Report	
	% Correct	N	% Correct	N	% Correct	N	% Correct	N
Year								
4	61	653	47	656	85	616	77	616
5	71	1872	60	1887	95	1844	83	1844
6	82	211	75	221	95	234	88	234

small in all cases except that 10-year-old students in non-metropolitan areas performed better on the task which required a telephone message to be taken.

Year Level. For both the 10- and 14-year-old groups only three year levels have been reported. Because of the small numbers of students in other year levels these were not reported.

From Table 8.13 it can be seen that in general the higher the year level the better the performance of students. The differences between year levels were greatest on the more difficult tasks. On the easiest task, *Write a Personal Message*, there was a substantial difference (10 per cent) between Years 4 and 5, but no difference between Years 5 and 6 (95 per cent). Reference to Table 8.14 which reports figures for the 14-year-old students, indicates that the percentage of correct responses tended to increase with higher year levels. The greatest difference between year levels was for the task *Write Directions*, which was the most difficult item. As might be expected, differences between year levels decreased as a higher percentage of the students achieved a correct response.

Table 8.14 Distribution of Correct Responses of 14-Year-Old Students on Four Tasks by Year Level

Task	Telephone Message		Personal Message		Write Directions		Write a Letter	
	% Correct	N	% Correct	N	% Correct	N	% Correct	N
Year								
8	88	664	95	627	42	627	82	627
9	95	1447	97	1697	57	1697	89	1697
10	94	225	98	203	69	203	92	203

Table 8.15 Distribution of Correct Responses on Three Writing Tasks by Original Students and Replacement Students

Students	Telephone Message		Personal Message		Write a Letter	
	% Correct	N	% Correct	N	% Correct	N
10-Year-Old Level						
Original	70	2533	92	2503	58	2561
Replacement	61	206	96	197	57	207
Total Sample	69	2739	93	2700	57	2768
14-Year-Old Level						
Original	92	2144	97	2259	87	2259
Replacement	93	273	97	283	89	283
Total Sample	92	2418	97	2542	87	2542

Performance in both age groups showed consistency when year levels were compared. The higher the year level the higher the percentage of successful students. Furthermore on more difficult items the disparity in the successful completion of tasks was greater between year levels.

Original and Replacement Schools and Students. It was considered possible that the results obtained for the samples of 10- and 14-year-old students could be affected by the replacement of schools and students declining to participate in the study. Accordingly an analysis has been made of the percentage distribution of correct responses of original and replacement schools and students, comparing these figures with the total samples for each of the writing tests which were approximately half the total sample. Table 8.15 shows that there was little variation between original and replacement students except at the 10-year-old level where there was a nine per cent difference in the task which required a telephone message to be taken. However, in no case did the percentage of correct responses of original students vary by more than one per cent from that of the total sample tested. This finding suggests that the incorporation of replacement students in the samples had a very small influence on results obtained for the writing tasks. From Table 8.16 which reports the distribution of correct responses on some tasks by students from original and replacement schools, it can be seen that there was little difference at the 14-year-old level between original and replacement schools. At the 10-year-old level differences were more marked for two tasks.

Table 8.16 Distribution of Correct Responses on Three Writing Tasks By Students at Original and Replacement Schools

Schools	Telephone Message % Correct	N	Personal Message % Correct	N	Write a Letter % Correct	N
10-Year-Old Level						
Original	70	2522	92	2597	57	2551
Replacement	59	217	96	103	70	217
Total Sample	69	2739	93	2700	57	2768
14-Year-Old Level						
Original	92	2342	96	2337	88	2337
Replacement	95	754	98	205	84	205
Total Sample	92	2418	97	2542	87	2542

It can be concluded that the effect of replacement schools and students on the total percentage of correct responses for writing tasks was minimal.

Summary

The tasks used in the writing tests varied in the demands they made on students. As wide a range of tasks as possible was selected for the tests but it was accepted that only a selection of the possible range and variety of tasks were able to be represented in the study.

As might be expected, a higher percentage of 14-year-old than 10-year-old students responded correctly to the same or similar criterion-referenced tasks. The more difficult the task the greater the difference between the two age levels. For instance, on a similar task which required a personal message to be written to a family member, at both age levels a high percentage of students responded correctly: at the 10-year-old level 93 per cent and at the 14-year-old level 97 per cent were successful. On the other hand on the task in which students had to write a letter requesting information and address an envelope, whilst 87 per cent of students at the 14-year-old level were successful only 57 per cent of 10-year-old students responded correctly. The most frequently omitted sub-task was the inclusion of a return address. However the part of the task which required an envelope to be addressed was done well by both age groups: 93 per cent of 10-year-old and 99 per cent of 14-year-old students were successful.

Differences in performance between 1975 and 1980 were small except for writing a personal message and some aspects of writing a letter for information in which 10-year-old students in 1980 performed better than those in 1975. At this age level students in 1975 and 1980 found difficulty with the task which required a telephone message to be taken: in 1980, 69 per cent and in 1975, 73 per cent responded correctly. On the other hand a larger proportion of 14-year-old students in both years completed this task successfully: in 1980, 92 per cent and in 1975, 89 per cent responded correctly. Although these differences were not significant within age levels, they were significant between age levels.

Of the tasks which were scored by global methods, writing opinions was the most difficult at the 14-year-old level and it also had the highest mean spelling error rate. Students produced fewer words for this task than for writing a story. At the 10-year-old level word production was greater for writing a story than for writing about experience. At this level fewest spelling errors were made in writing a report, a highly structured sequential task which drew on a limited and basic area of vocabulary.

There was a difference in performance for some background variables. Males performed less well than females at the 10-year-old level on all tasks. At the 14-year-old level males did less well than females on a task which required them to write a letter requesting information, and on all globally scored tasks. Year level made a difference to performance for both 10- and 14-year-old students, the higher the year level the higher the percentage of students completing tasks. For more difficult tasks the differences in performance between year levels were greater. There were small and variable differences in performance depending on the location of schools. The exception was that non-metropolitan students at the 10-year-old level performed significantly better on the task of taking a telephone message. The introduction of replacement schools and students had a minimal effect on the reported proportions of successful students.

CHAPTER NINE

PERFORMANCE ON THE NUMERATION TESTS

Method of Reporting Used

The Numeration Tests were described in some detail in Chapter Two. The results obtained by the two samples of students on these tests are reported in this chapter. The figures reported are the weighted results for the total Australian samples of 10- and 14-year-old students. In the main, two methods of reporting student performance have been adopted: reporting by individual item, or reporting by sub-tests which were formed by grouping items as suggested in Figure 2.3 into categories based on the three dimensions of content, ability and usage plus sub-tests for each of the four arithmetic operations. For both methods of reporting the proportion of successful students has been given.

Where differences between proportions of successful students on individual items are considered, Tables 3.4 and 3.5 should be consulted for an indication of what constituted a significant difference. When differences between proportions of successful students on sub-tests (that is, those who achieved the 80 per cent mastery criterion) are considered, two cautions need to be exercised. First, the actual mastery levels set differ somewhat from 80 per cent according to the number of items in a sub-test and secondly, the mastery level required depends upon the relative proportions of multiple-choice and constructed-response items in each sub-test.

The proportions of successful students on each of the sub-tests and on the total test are reported in this section. Where it seems of particular interest the results for individual items are also given together with a reproduction of the actual item from the test. Details of results for every item have been provided in Appendix Three to this report. The results on thirteen sub-tests and the total test for the 10-year-old students are discussed first followed by the results on fourteen sub-tests, some individual items and the total test for the 14-year-old students. These results are shown in Tables 9.1 and 9.2 for the 10- and the 14-year-old students respectively. The classifications of items into sub-tests based on the three dimensions of content, ability and usage were discussed in Chapter 2 and were set out in Tables 2.8 and 2.9 for the 10-year-old and the 14-year-old students respectively. Sub-test and the total test reliabilities have been reported in Appendix Five.

Table 9.1 Numeration Test for 10-Year-Old Australian Students: Proportions of Successful Students on Each Sub-test and the Total Test

Dimension: Sub-test	No. of Items	No. of Items Required	Successful Students %
Content:			
Numbers (Whole & Fraction)	14	10	84
Measurement-General	6	4	78
Measurement-Money	5	3	95
Measurement-Time	7	5	72
Ability:			
Knowledge	18	13	85
Computation	4	2	67
Application	13	9	84
Usage:			
Community	21	16	74
Classroom	14	10	84
Operation:			
Addition	5	3	98
Subtraction	5	3	86
Multiplication	4	2	94
Division	4	2	87
Total Test	35	26	78

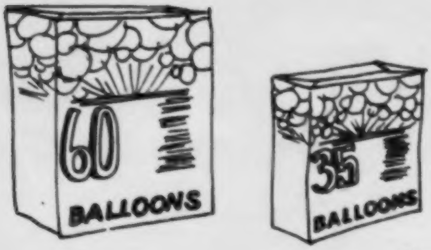
Results for the 10-Year-Old Students

Content Dimension

There were four categories on the content dimension: Numbers (Whole and Fraction), Measurement-General, Measurement-Money and Measurement-Time. The numbers of items making up each sub-test are shown in Table 9.1 together with the number of items required to be correct if the student was to be considered successful on the sub-test (that is, the 80 per cent mastery criterion level), and the proportions of successful students for each sub-test. The actual item numbers for all sub-tests have been given in Chapter Two, Table 2.8.

Numbers (Whole and Fraction). This sub-test consisted of 14 items of which at least 10 items correct were required. This result was achieved by 84 per cent of the 10-year-old students. The items making up this sub-test included one item on place value, two items involving fractions, eight items requiring knowledge of and computation with whole numbers, and three items where the application of whole-number skills to problems was necessary. One example of this latter type of item is shown in Figure 9.1. This item was answered correctly by 94 per cent of the students.

21.



How many balloons are there altogether?

Answer ____________ balloons

Figure 9.1 Item Requiring the Application of Whole Number Skills to Problem Solving.

3. John's dad is a tall man. His height is about

☐ 1 metre
☐ 2 metres
☐ 4 metres
☐ 6 metres

4. Which of these is usually about 10 metres wide?

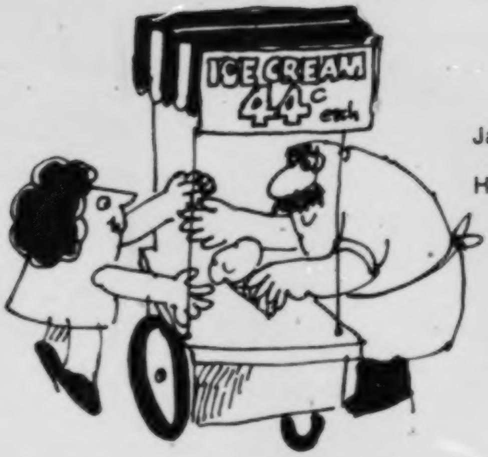
☐ a plate
☐ a door
☐ a car
☐ a road

Figure 9.2 Two Items Involving Estimation of Metric Lengths.

Measurement-General. Students were required to respond correctly to at least four of the six items in this sub-test and 78 per cent of the 10-year-olds did so. Two items required some estimation of metric length, two items involved reading charts, one item was concerned with the division of two volumes, and one item with the measurement of length. Where units were involved, all were in metrics. Two items involving estimation of metric lengths provided a startling contrast in terms of the results observed. Whereas only 26 per cent of the 10-year-olds could correctly select the most appropriate height of a tall man from four alternatives all in metres, 82 per cent selected which of four objects was approximately 10 metres wide. These contrasting items are shown in Figure 9.2. It is perhaps of interest that the former item was by far the most difficult item in the numeration test for the 10-year-old students.

Measurement-Money. At least three of the five items were correctly answered by 95 per cent of students who thus met the requirement for this sub-test. This high level of performance was reached on items concerning the use of the four operations with money in problems, and one item requiring the addition of six coins totalling 88 cents. The most difficult item in this sub-test was correctly answered by 71 per cent of the students. This item is shown in Figure 9.3.

29.



Jan gave 60 cents for an ice cream.

How much change should she get?

Answer ______________ cents

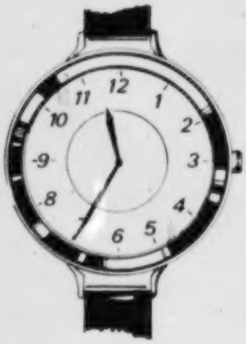
Figure 9.3 Money Item Requiring Subtraction with Amounts Less than One Dollar.

Measurement-Time. Students were required to answer correctly five of the seven items constituting this sub-test and 72 per cent were able to reach this level of performance. Two of the items required only the reading and recording of the time shown on a watch-face, one item required reading a watch-face and a digital clock and making a comparison of the two times, two items involved reading a calendar, one item the calculation of a time 20 minutes after a given time, and one item finding a time and the corresponding entry in a television program. The two items requiring the reading and recording of times on a watch-face are shown in Figure 9.4. It is of interest that 71 per cent of the 10-year-olds were able to do this when the watch had Arabic numerals on the face, while only 62 per cent were correct when the watch-face did not have numerals. As shown subsequently in this chapter (see Table 9.11), the 71 per cent of 10-year-olds able to read the watch-face with Arabic numerals was identical with the result obtained for the same item in the 1975 ASSP study.

Ability Dimension

Three categories - knowledge, computation and application - were identified on the ability dimension for the numeration tests. A sub-test was formed to assess student performance on each category and these results are now reported.

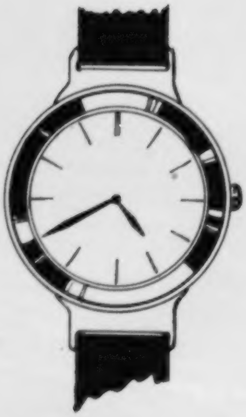
9.



What time is shown by the watch?

Answer ____________________

33.



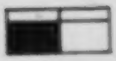
Write the time shown by this watch.

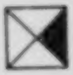
Answer ____________________

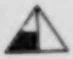
Figure 9.4 Reading Time from Watch-faces With and Without Numerals.

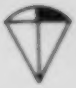
Knowledge. For the 10-year-old students knowledge was considered to be the most important category on the ability dimension and more than half the items tested knowledge. There were 18 items comprising this sub-test and students were required to answer at least 13 items correctly. This level of performance was reached by 85 per cent of the 10-year-olds. Knowledge required included counting and ordering, place value, recognition of fractions and metric measurements, knowledge of tables, telling the time and reading calendars. The two items involving the recognition of fractions were different in that one item required the recognition of one quarter of a cube formed by diagonals from other fractions of a rectangle, a triangle and an irregular figure, and the other item required the recognition of three quarters from other fractions including one quarter and one half all based on segments of a circle. The actual items are shown in Figure 9.5. The proportions of students correct were almost identical at 88 per cent and 89 per cent respectively.

2. What shape has $\frac{1}{4}$ shaded?

☐ 

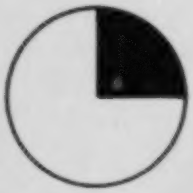
☐ 

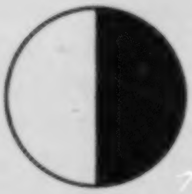
☐ 

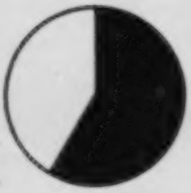
☐ 

5. Faye coloured $\frac{3}{4}$ of her circle black.

Put a cross under Faye's circle.


______


______


______

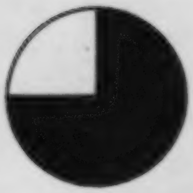

______

Figure 9.5 Two items Involving the Recognition of Fractions

Computation. There were only four items in this category with most emphasis being directed at the knowledge and application categories of ability. The items were one each of the four operations of the type '38+19' and all are shown in Table 9.8. Although the cutting score that minimized errors of classification for this sub-test was only two out of the four items correct, this level of performance or higher was reached by only 67 per cent of the 10-year-old students.

Application. Application was also seen as an important category on the ability dimension and more than a third of the items tested application. Of the 13 items in this sub-test, students needed to respond correctly to at least 9 items and this level was reached by 84 per cent of the 10-year-olds.

The items included use of the four operations in problem settings (frequently involving money), plus some counting, reading of charts and time comparison. Some of the items in this sub-test were similar with respect to the arithmetic process required though different in context from some items in the computation sub-test where they were not in problem settings. It is of interest that student performance was similar on items involving similar arithmetic processes in different contexts. For example, when asked to obtain the total number of cards in 4 boxes of 15, most students chose to add four 15s rather than to multiply 15 by 4, and 89 per cent were successful. This compares with the 88 per cent of students who correctly added 38 and 19 which was one item of the computation sub-test.

Usage Dimension

Two categories were identified on the usage dimension. These were community usage in which numerical skills necessary in everyday life were assessed in appropriate contexts and classroom usage in which similar numerical skills considered necessary for further school work with number were tested in a manner similar to that used in textbooks. Results obtained by the 10-year-old students on the community and classroom sub-tests formed from the appropriate items are now described and discussed.

Community Usage. The community usage sub-test comprised 21 items, at least 16 of which were to be answered correctly to meet the required level of performance. This level was reached or surpassed by 74 per cent of students. This sub-test included all five money items together with the three counting and ordering items and items requiring use of one or more of the four arithmetic operations. As stated above each item was placed in an out-of-school context which was intended to be familiar to students.

Classroom Usage. Of the 14 items forming the classroom usage sub-test, 10 items were required, and 84 per cent of the 10-year-olds reached this level of performance or higher. Most items involved use of the four operations but other items were concerned with knowledge of place value, recognition of fractions, interpretation of charts and simple measurement.

The Four Operations

There were small numbers of items which could be answered by use of one of the four operations, and which, it was judged, the vast majority of students would attempt to answer by using one of addition, subtraction, multiplication or division. The results for these four sub-tests are also shown in Table 9.1. Very high proportions of the 10-year-olds reached the required performance levels for addition (98 per cent) and multiplication (94 per cent) with only slightly lower proportions for subtraction (86 per cent) and multiplication (87 per cent). With the small numbers of items involved, particularly for multiplication and division, these results are very dependent upon the actual individual items used in each case. However it is considered that the results obtained are indicative of the relative proportions of students aged 10 years who can perform the four operations.

Counting and Ordering

There were only three items which assessed counting and ordering and these are shown in Figure 9.6. Item 6 which required the selection of 172 as the highest numeral was answered correctly by 97 per cent of the 10-year-old students, item 7 requiring the identification of 'fourth' by 96 per cent, and item 8 counting 14 people by 99 per cent. Thus it can be seen that almost all 10-year-olds were able to handle the tasks of counting and ordering successfully.

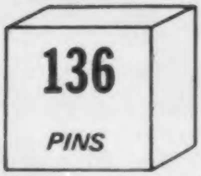
Total Test Results

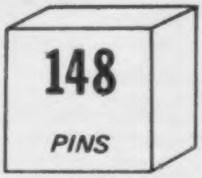
Although the major interest in student results for this study centred on performance on specific tasks and thus on items and sub-test scores, the total test result does provide a useful overview of performance on the numeration test. There were 35 items in the test and the criterion level of performance was reached by the 78 per cent of students who obtained at least 26 items correct.

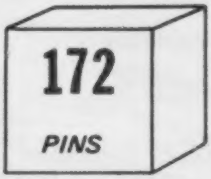
Figure 9.7 uses the proportion of students who succeeded on the test taken as a whole as a basis for comparing the proportions of successful students on each of the individual sub-tests. It will be noted that only on three sub-tests was the proportion of successful students lower than that for the total test. The sub-tests where student performance was lower were the content sub-test on time, the ability sub-test concerned with computation, and the usage sub-test assessing community-based numerical skills. When considering Figure 9.7 it should be remembered that the proportion of items required to be correct did vary between sub-tests as a result of varying test lengths and types of item used which permitted different possibilities of guessing the correct response. Consequently the figure provides no more than a guide to the students' relative performances on the various sub-tests.

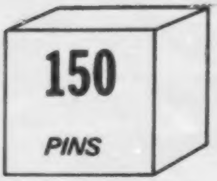
One other aspect of the whole test should be mentioned. It was intended that all students should be able to complete the test in the 30 minutes allowed.

6. Put a cross under the box with the most pins in it.



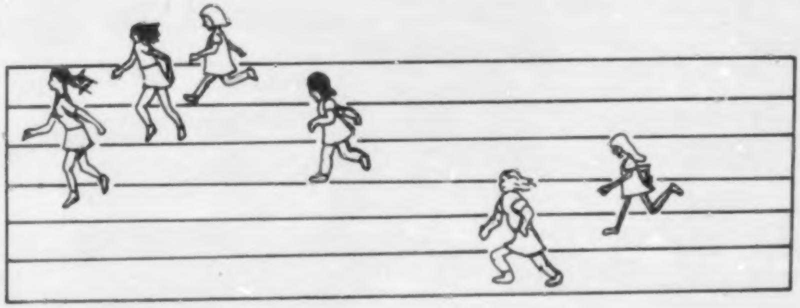






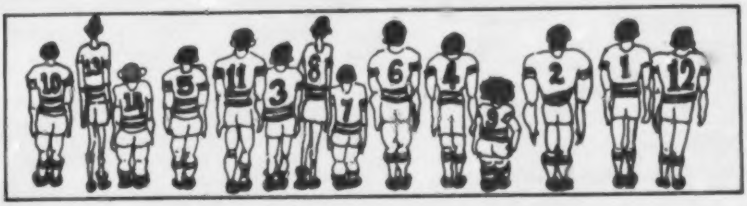
__________ __________ __________ __________

7.



Put a cross on the girl who is running fourth in the race.

8.



How many people are in this picture?

Answer ____________ people

Figure 9.6 Items Testing Counting and Ordering.

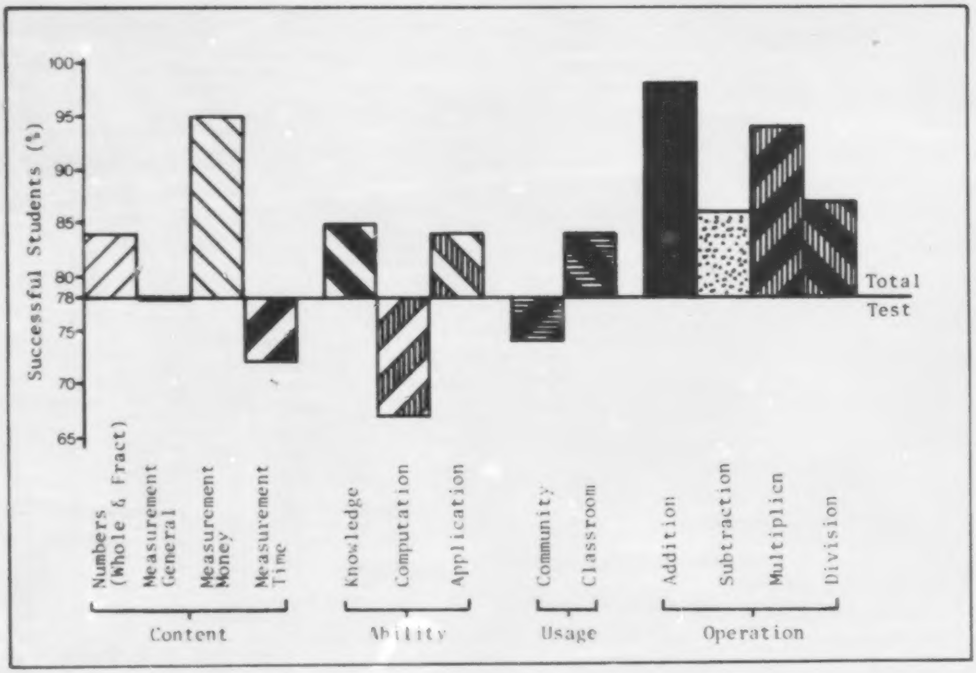


Figure 9.7 Proportions of Successful 10-Year-Old Students on the Numeration Sub-tests Compared with the Total Test.

Inspection of Table A3.2 in Appendix Three shows that only two per cent of the 10-year-old students did not respond to the last item in the test and that the proportion of students not answering an item tended to be related to the difficulty level of the item concerned. It would seem that all, or almost all, students did have sufficient time to attempt all items.

Results for the 14-Year-Old Students

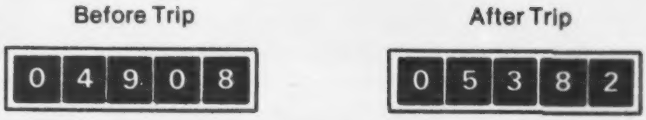
The results obtained by the 14-year-old students on the fourteen numeration sub-tests are now described. These results, together with the number of items in each test and the number of items required for a student to be considered to have reached the 80 per cent mastery criterion level, are shown in Table 9.2. As was the case with the 10-year-old students, the sub-tests have been grouped according to three dimensions - content, ability and usage - and are reported in that order followed by the sub-tests concerned with the four arithmetic operations. The actual item numbers for all sub-tests have been shown in Chapter Two, Table 2.9.

Table 9.2 Numeration Test for 14-Year-Old Australian Students: Proportions of Successful Students on Each Sub-test and the Total Test

Dimension: Sub-test	No. of Items	No. of Items Required	Successful Students %
Content:			
Whole Numbers	11	8	95
Fractions & Decimals	7	5	80
Measurement-General	7	5	84
Measurement-Money	5	3	95
Measurement-Time	5	3	96
Ability:			
Knowledge	11	8	96
Computation	7	5	92
Application	17	12	83
Usage:			
Community	16	11	89
Classroom	19	14	91
Operation:			
Addition	4	2	99
Subtraction	8	5	92
Multiplication	5	3	93
Division	5	3	93
Total Test	35	26	86

18.

Kilometre Readings



What distance was covered on this trip?

Answer_________________km

Figure 9.8 Whole Number Item Involving Application of Subtraction with a Community-based Usage.

Content Dimension

There were five sub-tests on the content dimension: two were concerned with whole numbers and with fractions and decimals, and the remaining three sub-tests in the measurement area were the same as for the test for the 10-year-olds, namely measurement-general, measurement-money and measurement-time.

Whole Numbers. Of the 11 items in this sub-test, at least 8 items were required to be correct and this level of performance was achieved by 95 per cent of the 14-year-old students. The items were largely those concerned with knowledge and computation with the four operations in a classroom context although one place value and one community-based item were included. This latter item, shown in Figure 9.8, was answered correctly by 77 per cent of the 14-year-old students. This item required the subtraction of one four-digit numeral from another and was found to be more difficult than other items requiring subtraction.

Fractions and Decimals. There were seven items in this sub-test and students were required to answer at least five items correctly. This level of performance was achieved by 80 per cent of the 14-year-old students. Two items involved working with common fractions and the remaining five items involved decimal fractions, sometimes in a classroom context and otherwise in a community-based problem format. The item found to be the most difficult overall by the 14-year-olds was contained in this sub-test. This item, shown in Figure 9.9, required students to change halves, quarters and eighths to

6. **Which fraction is closest in size to $\frac{3}{16}$?**

☐ $\frac{1}{16}$

☐ $\frac{1}{4}$

☐ $\frac{3}{8}$

☐ $\frac{7}{16}$

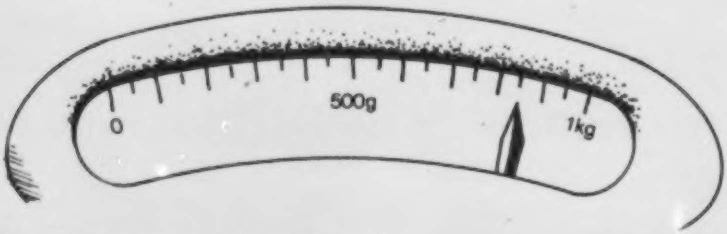
☐ $\frac{1}{2}$

Figure 9.9 Most Difficult Item on the Numeration Test for 14-Year-Old Students.

sixteenths to find the fraction closest to 3/16 and was answered correctly by only 58 per cent of the 14-year-olds.

Measurement-General. This sub-test also consisted of seven items and 84 per cent of the 14-year-olds answered correctly five or more items to reach the required level of performance. Five of the seven items required the ability to apply numerical skills to the solving of problems involving graphs, maps, distance, area and volume, and the remaining two items were concerned with the measurement of length and of mass. The item involving the reading of a scale in a community context is shown in Figure 9.10. This item was correctly answered by 83 per cent of 14-year-old students.

28.



What is the reading shown above?

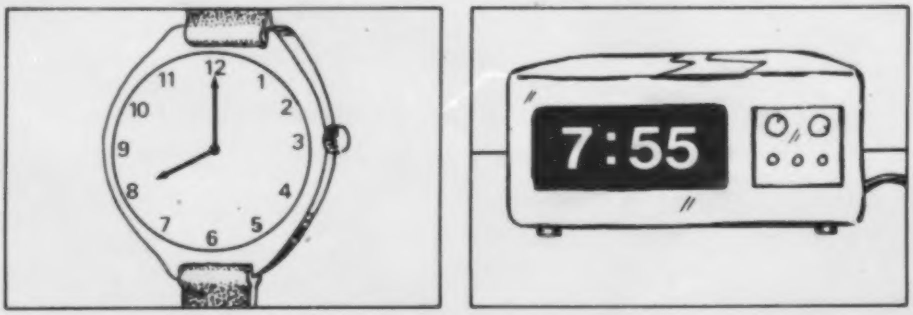
Answer ______________ grams

Figure 9.10 Item Involving Measurement of Mass in a Community Context.

Measurement-Money. Students were required to answer three or more of the five items of this sub-test correctly and this result was achieved by 95 per cent of the 14-year-olds. Four of the five items required the application of the four operations and the other item involved only knowledge of how to write the numeral 162 in words on a facsimile of a bank withdrawal form. Spelling was not a criterion as long as the amount written could be clearly identified as 162, and this item was answered correctly by 94 per cent of the 14-year-old students. The item where students had to calculate 10 per cent interest on $900 was the least well-answered in this sub-test in that only 78 per cent of the students gave the correct answer. However it will be noted that this was a significantly higher proportion correct than that obtained in the 1975 ASSP project when performances of the 14-year-old students on identical items are compared later in this chapter.

Measurement-Time. There were also five items in this sub-test and 96 per cent of the 14-year-old students obtained three or more items correct and thus met the required level of performance. Two items required knowledge of how to read a calendar and a watch face and the other three items involved application of the four operations with time. On no item in this sub-test did the proportion of students correct fall below 85 per cent. One item of particular interest was that shown in Figure 9.11 which required the calculation of a time difference between a circular watch-face and a digital clock. This item was answered correctly by 95 per cent of the 14-year-old students in the sample.

25.



Pam checked her watch against a digital clock.

What is the difference between the two times?

Answer ____________ minutes

Figure 9.11. Item Involving Difference Between Times Shown on a Circular Watch-face and a Digital Clock.

1. In the number 3251, what does the 5 stand for?

- ☐ 5 thousands
- ☐ 5 hundreds
- ☐ 5 tens
- ☐ 5 ones

2. In the number 23·64, what does the 6 stand for?

- ☐ 6 tens
- ☐ 6 ones
- ☐ 6 tenths
- ☐ 6 hundredths

Figure 9.12. Items Involving Knowledge of Place Value.

Ability Dimension

The three categories of knowledge, computation and application reported for the 10-year-old students were also applied at the 14-year-old level and the results on these sub-tests are now reported.

Knowledge. There were 11 items in this sub-test and students were required to answer at least 8 items correctly to demonstrate the required level of competence in this area. This was achieved by 96 per cent of the 14-year-olds. Items required knowledge of place value, expressing numerals in words, tables, and the measurement of length, mass and time. Student performance on these items was uniformly high with 81 per cent responding correctly to the most difficult item in the sub-test. This item required knowledge of the place value 'tenths'. Another item involving knowledge of the place value 'tens' was correctly answered by 97 per cent of the 14-year-olds. Both of these items are shown in Figure 9.12.

Computation. Students were required to answer correctly at least five of the seven items making up the computation sub-test and this level of performance was achieved by 92 per cent of the 14-year-olds. Five items involved computation with whole numbers and two with decimal fractions.

Put a cross in the correct box.

35. **1240 ÷ 29 is about**

☐ 20
☐ 30
☐ 40
☐ 50
☐ 60

Figure 9.13 Item Requiring Computation with Whole Numbers Involving Four Digits.

Performance on the decimal fraction items was comparable with that for four of the five whole number items where mostly two but no more than three digits were involved. The other whole number item which was found to be more difficult involved four digits and required the students to select from five alternatives the approximate answer to an item requiring the operation of division. The item, shown in Figure 9.13, was answered correctly by 75 per cent of the 14-year-olds.

Application. Whereas emphasis at the 10-year-old level tended towards knowledge on the ability dimension, an emphasis on application seemed more appropriate for the 14-year-old students. Consequently there were 17 items in this sub-test and the correct answer was required to 12 or more items. This level of performance was achieved by 83 per cent of the 14-year-old students. The items included use of the four operations (mostly involving money or fractions), reading of graphs, calculation of time, and volume. Apart from the fraction item previously shown in Figure 9.9, the two most difficult items in this sub-test were items 23 and 33 and these items are shown in Figure 9.14. Item 23 required the subtraction of two decimal fractions and item 33 required comparison of two volumes and these items were answered correctly by 67 per cent and 63 per cent of students respectively.

Usage Dimension

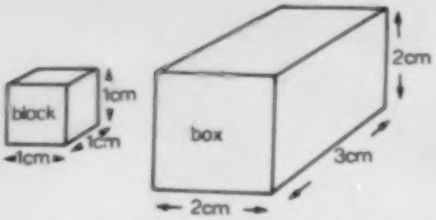
The same two categories of community and classroom usage as described for the test for the 10-year-old students were used to form sub-tests at the 14-year-old level. Results on these sub-tests are now given.

23.

Sue and Jill ran a 70 metre race.
Sue's time was 10·8 seconds.
Jill's time was 12·4 seconds.

How many seconds faster was Sue's time? Answer ____________ seconds

33.



How many blocks would fill the box?

Answer ______________ blocks

Figure 9.14. Two Difficult Items in the Application Sub-test.

Community Usage. Of the 16 items comprising the community usage sub-test 11 correct items were required and 89 per cent of the 14-year-olds reached this level or higher. This sub-test consisted of most of the money items, all the time items, interpreting a map scale, calculating time and distance and use of the four operations in everyday contexts.

Classroom Usage. Of the 19 items in the classroom usage sub-test, 14 items were required to be correct and this level of performance was achieved by 91 per cent of the 14-year-old students. Items required knowledge of place value and of tables, computation involving the four operations with whole numbers and fractions, measurement of length, area and volume, and the reading of graphs.

The Four Operations

As was the case for the 10-year-old students, there were small numbers of items which required knowledge, computation or application of one of the four arithmetic operations of addition, subtraction, multiplication or division. These items were formed into four sub-tests, the results of which are also shown in Table 9.2 where it can be observed that 99 per cent of the 14-year-olds reached the required level for addition, and slightly smaller proportions were successful for subtraction (92 per cent) and for multiplication and division (both 93 per cent). Although the performance level that minimized classification errors of students for the addition sub-test was only two out of the four items, the result would have been

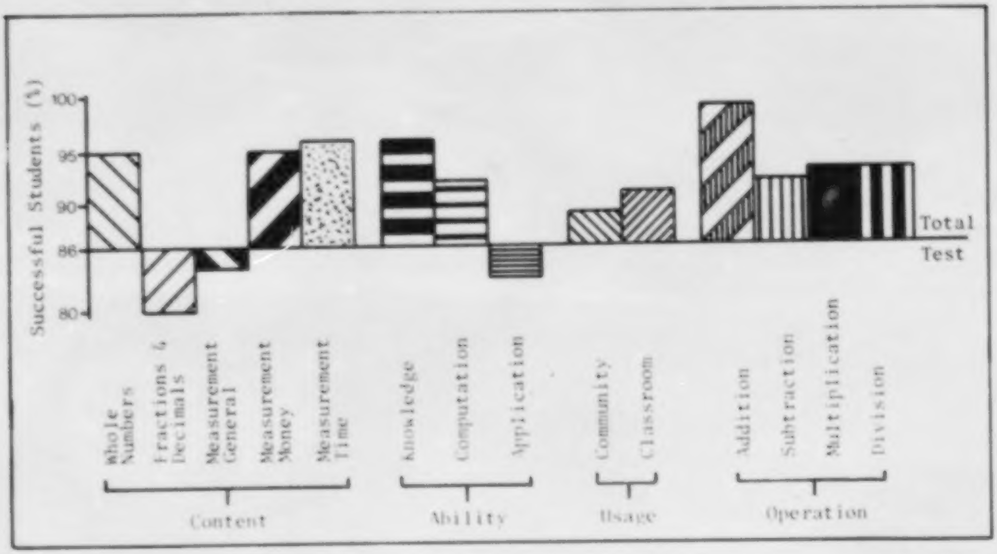


Figure 9.15 Proportions of Successful 14-Year-Old Students on the Numeration Sub-tests Compared with the Total Test.

little different if the requirement had been three items correct. In that case 96 per cent of the 14-year-old students would have been considered successful on the addition sub-test.

Total Test Results

To provide an overview of the numeration test results for the 14-year-old students, the proportion of students reaching or surpassing the required level of 26 out of the 35 items comprising the test was calculated. As shown in Table 9.2, 86 per cent of the students reached this performance level or higher. In fact almost half the students had a score of 32 or higher on the total test indicating that they were successful on approximately 90 per cent of the items and thus exceeded comfortably the 80 per cent criterion level set throughout this study. As stated for the 10-year-old students, consideration of the total test score is not intended as a substitute for the more detailed results on specific tasks, sub-tests and items.

The proportions of successful students on each of the sub-tests were compared using the total test result of 86 per cent as the basis of comparison. The comparisons are illustrated in Figure 9.15 which shows that, compared with the total test result, student performance was lower for only three of the fourteen sub-tests. For two of the sub-tests (the content sub-test measurement-general, and the application ability sub-test) the difference was marginal but for the content sub-test concerning fractions and decimals the difference was greater. The proportions of successful students for the remaining eleven sub-tests were generally above 90 per cent and so were considerably higher than for the total test. Again it should be stressed that these comparisons, based on different test lengths and different item formats, should only be used as a guide to the relative levels of student performance on the numeration tasks tested.

The proportions of students who were correct, incorrect or did not answer each item of this test are shown in Table A3. in Appendix Three. Inspection of the 'no attempt' column suggests that almost all students had sufficient time to complete the test. For most items no more than two per cent of the students failed to respond and this was true of the last item in the test. There were only three items for which more than two per cent of students made no attempt: these were items 18, 31 and 34. From comments made by supervising teachers, almost all students had ample time to complete the numeration test.

Results by Background Variables

Comparison of student results on the numeration tests by a small number of background variables are presented in this section. Results for male and female students are described in some detail and results by location (metropolitan/non-metropolitan) and year level are described briefly. Finally results for students originally selected in the samples are compared with the results for students who replaced other students either because a school did not take part in the program or because a student was absent on the day of testing. Some concern had been expressed that replacement students might distort the overall results obtained.

Results by Sex

10-Year-Olds. When individual items were considered, there were nine items at the 10-year-old level where there was a significant difference in performance between male and female students. The degree of difference required for significance is shown in Chapter Three (Table 3.7), but for most of the items in the numeration test for the 10-year-old students a difference of approximately 5 per cent was necessary. It will be observed in Table 9.3 where a brief description of each of the nine items for which there was a

Table 9.3 Percentages of 10-Year-Old Male and Female Students Responding Correctly to Items Exhibiting a Significant Difference in Performance Between the Sexes

Item No.	Description of Task Tested	% Correct	
		Male	Female
3	Estimate in metres a length less than 10m	20	32
12	Recall multiplication facts for 2 numerals each less than 10	85	90
14	Read a calendar	69	79
16	Add two numerals each of 2 digits	86	90
17	Subtract two numerals each of 2 digits	69	79
18	Multiply a 2-digit numeral by a 1-digit numeral	55	66
27	Reading a pictograph	89	94
30	Subtract whole numbers to solve problems	75	80
32	Divide whole numbers to solve problems	72	66

difference between the sexes is given together with the proportions of males and females correct on the item, that for eight of the nine items there were significantly higher proportions of female students who obtained the correct answers. The only item on which the males had higher performance required the division of 150 by 10 to solve a problem. There would seem to be no obvious reason other than chance for the reversal of results for this item.

When sub-test results were considered there were differences on four of the thirteen sub-tests at the 10-year-old level, all of which favoured the female students. As can be seen in Table 9.4, these were the content sub-test concerning numbers, the ability sub-test involving computation, the classroom usage sub-test and the operation sub-test on subtraction. For all the other sub-tests and the total test, differences in the proportions of male and female students reaching the mastery criterion level were small.

14-Year-Olds. Differences in performance between the sexes were generally smaller for the 14-year-olds than was the case for the 10-year-olds. At the 14-year-old level, there were eight items where student performance differed between the sexes (see Table 9.5). However in this case the results on six of the items favoured the males and the other two results favoured the females. The largest difference was for the item involving the volumes of

Table 9.4 Proportions of 10-Year-Old and 14-Year-Old Male and Female Students Successful on the Numeration Sub-tests and Total Test

Dimension:	Successful Students			
	10-Year-Old		14-Year-Old	
Sub-test	Male %	Female %	Male %	Female %
Content:				
Whole Numbers	-	-	94	96
Fractions & Decimals	-	-	81	80
Numbers (Whole & Fraction)	81	86	-	-
Measurement-General	79	78	88	80
Measurement-Money	95	95	94	95
Measurement-Time	72	73	96	96
Ability:				
Knowledge	84	86	95	96
Computation	81	89	92	93
Application	84	84	84	81
Usage:				
Community	74	74	89	88
Classroom	81	87	91	90
Operation:				
Addition	98	99	99	99
Subtraction	83	88	92	92
Multiplication	93	94	93	94
Division	86	87	93	93
Total Test	77	80	87	86

two cuboids for which 14 per cent more males than females gave the correct answer. The two items on which the females performed better than the males were a computation question requiring subtraction and reading a calendar.

Sub-test results comparing results by sex for the 14-year-old students are shown with the results for the 10-year-olds in Table 9.4. The only sub-test for which there was a difference in performance between the sexes was the measurement-general sub-test on the content dimension. This sub-test of seven items required the reading of graphs, maps and scales and it was found that 88 per cent of the males reached the required level of performance

Table 9.5 Percentages of 14-Year-Old Male and Female Students Responding Correctly to Items Exhibiting a Significant Difference in Performance Between the Sexes

Item No.	Description of Task Tested	% Correct	
		Male	Female
5	Read a map to obtain distance	87	82
6	Add and subtract common fractions	62	54
12	Measure the length of an object in mm	94	89
14	Subtract two numerals each of 2 digits	90	94
17	Read a calendar	87	93
21	Read time on a watch-face without Arabic numerals	90	86
28	Read mass indicated on scales in gm	87	79
33	Compare volumes of cuboids	69	55

compared with 80 per cent of the female students. For all other sub-tests and the total test, the performances of the male and female students were very similar.

Results by School Location

When the proportions of students who succeeded on the total numeration test by reaching the 80 per cent criterion level were considered, it was found that, whereas 81 per cent of the 10-year-old students from metropolitan schools had reached the criterion level or higher, only 75 per cent of the non-metropolitan students had reached this level. For the 14-year-old students, the proportion of successful students was 86 per cent in each case. The reasons for the difference found for the 10-year-old students are not clear but are possibly related to other background factors for which information was not collected. In any case the small difference favouring the metropolitan students at the lower age level was not replicated by the 14-year-old students.

When school location was considered in relation to the various numeration sub-tests the patterns were the same as for the total tests. At the 10-year-old level there were generally small differences favouring the metropolitan students and at the 14-year-old level there were no differences in the proportions of successful students from metropolitan and non-metropolitan locations. Sub-test results have not been reported in detail as they would be simply repetitious.

Table 9.6 Year Level of Students: Numbers and Proportions of Samples and Proportions of Successful Students on the Total Numeration Test

10-Year-Olds				14-Year-Olds			
Year Level	N	% Sample	% Success	Year Level	N	% Sample	% Success
4	1274	23	57	8	1341	26	74
5	3733	68	83	9	3228	63	91
6	455	8	94	10	436	9	96
Other	9[a]	-	-	Other	97[b]	2	-

[a] Four students were in Year 3 and year level was unavailable for five students.

[b] Made up of 37 students in Year 7 and 60 students for whom year level was not available.

Results by Year Level

10-Year-Olds. The numbers and percentages of 10-year-old students in Years 4 to 6 are shown in Table 9.6 where it will be noted that 68 per cent were in Year 5, 23 per cent in Year 4 and 8 per cent in Year 6. As might be expected, more students in higher year levels were able to complete the numeration tasks successfully compared with students in lower year levels. The proportions of successful students for the total numeration test are also shown in Table 9.6. Results on the sub-tests were similar with even larger differences for the time and the division sub-tests, and smaller differences for the sub-tests concerned with money, addition and multiplication.

14-Year-Olds. The majority of 14-year-old students were in Year 9 as shown in Table 9.6, however more than a quarter were in Year 8 and almost 10 per cent in Year 10. Once again more students in higher year levels reached the levels required for the sub-tests, although the differences were smaller than had been the case for the 10-year-old students. The largest differences in performance were found for the sub-tests concerned with fractions and decimals, measurement-general and application where approximately 20 per cent more of the Year 10 students were successful when compared with students in Year 8.

Replacement Schools and Replacement Students

When a school drawn in one of the samples declined to take part in this study, where practicable it was replaced by a school as much like the original school as possible. Details of replacement procedures for schools and for

Table 9.7 Replacement Schools and Students: Numbers and Proportions of Successful Students on the Total Numeration Test

	Schools			Students		
	N	%	% Success	N	%	% Success
10-Year-Olds						
Original	5150	94	78	5068	93	78
Replacement	323	6	84	404	7	82
Total Sample	5472	100	78	5472	100	78
14-Year-Olds						
Original	4822	94	86	4536	89	86
Replacement	281	6	84	567	11	86
Total Sample	5103	100	86	5103	100	86

students have been given in Chapter Three. The numbers and proportions of students from original and from replacement schools have been shown in Table 9.7, together with the proportions of successful students in each category. It will be noted that only 6 per cent of the overall samples consisted of students from replacement schools. There was no difference at either age level between the proportions of successful students for the total sample and for the originally-drawn sample. The differences in the proportions of successful replacement students are small when the relatively small numbers of replacement students are considered.

When a student drawn in the random sample of up to 25 students to be tested at a school taking part in the testing program was not present, he or she was replaced by another student who had also been drawn at random from all students of the appropriate age at that school. In the case of the sample of 10-year-olds, 7 per cent of the students in the achieved sample were replacements for other students; for the sample of 14-year-olds, 11 per cent were replacements for absent students. Table 9.7 also illustrates the numbers and proportions of replacement students and the proportions of successful students in the original, replacement and total samples. Again it will be noted that there was no difference between the original and total samples in the proportions of successful students for the total numeration test. The small difference in performance between the original and replacement students at the 10-year-old level was not significant given the relatively small number of replacement students.

Comparisons of Student Performance in 1975 and 1980

The Four Operations

Two major comparisons were made between performances on the numeration tests in the 1975 and the 1980 testing programs. The first comparison was that between eight basic items concerned with the four operations: four items involved only knowledge of tables, and four items each required the use of one of the four arithmetic operations, the operation required being specifically stated. These items consisted of identical arithmetic requirements on both occasions although the format varied. In 1975 two cues to the operation required were given in each case whereas only one cue was given in 1980. For example, the addition table item was presented in 1975 as:

Add $9 + 6$

and was presented in 1980 simply as:

$9 + 6$

In both cases a horizontal format was used. When items were field tested both presentations were used and no difference was found in the performances of students which could be attributed to the variation in presentation.

The second major comparison made was between thirteen items that were identical in all respects in both the 1975 and the 1980 test administrations except that their position in the test varied. In 1980 some items were later in the test and other items were earlier than had been the case in 1975.

It will be noted in Table 9.8 that the 10-year-old students performed considerably better in 1980 than in 1975 on six of the eight items in this group. Performance on the other two items was essentially the same on both occasions. The knowledge of basic tables had improved for each of the operations of addition, subtraction, multiplication and division.

The 14-year-old students had a higher level of performance on three of the eight items in 1980 and their performance on the other five items did not differ significantly between 1975 and 1980. However as very high percentages had been correct in 1975 on most of these items, there was a limit to the upward change that could have been anticipated. It was notable that in no case, at either age level, was performance on these basic items lower in 1980 than it had been in 1975.

Other Link Items

A group of thirteen items representing the range of items across the test as a whole was selected at each age level to form the common item link between

153

Table 9.8 Comparisons of the Percentages of Students Correct on Eight Basic Items Between 1975 and 1980

Item	10-Year-Olds		14-Year-Olds	
	1975 %	1980 %	1975 %	1980 %
Basic Tables				
9 + 6	91	97[a]	99	99
17 - 9	87	92[a]	96	97
7 x 6	83	87[a]	96	97
56 ÷ 7	73	80[a]	92	95[a]
Four Operations				
38 + 19	83	88[a]	94	94
36 - 19	74	74	90	92
38 x 9	55	60[a]	80	85[a]
125 ÷ 5	65	66	90	93[a]

[a] The differences in the proportions of students obtaining the correct answer are significant in that each difference is greater than the sum of two standard errors for each of the 1975 and the 1980 test administrations.

the two test administrations in 1975 and 1980. The link items differed between the two age levels. The distributions of students for each sub-test formed from the thirteen items are shown in Table 9.9 in the form of cumulative percentages of students correctly responding to items. The mastery level required in each case for these sub-tests was nine items correct. This level of performance was attained by 85 per cent of the 10-year-old students in 1975 and again in 1980. In fact the proportions of students across the distribution were remarkably consistent at this age level, further indicating that the level of performance on these items did not differ between 1975 and 1980.

In the case of the 14-year-old students, 79 per cent and 83 per cent obtained at least nine items correct for this sub-test in 1975 and 1980 respectively. After correction for the complex sample design used, application of Zubin's nomograph to these percentages indicated that this difference was significant at the 5 per cent confidence level (effective sample sizes were $n_1=1785$, $n_2=1458$ for a value of $\mathrm{deff}^{\frac{1}{2}}=3.5$)(Oppenheim, 1968:287-8).

The link items were then considered individually in order to determine for the 10-year-olds whether the overall result of no significant difference

Table 9.9 Comparisons of Cumulative Percentages of Students Correct on Link Items Between 1975 and 1980.

No. of Items Correct[a]	10-Year-Olds 1975 Cum.%	10-Year-Olds 1980 Cum.%	14-Year-Olds 1975 Cum.%	14-Year-Olds 1980 Cum.%
0	0.2	0.1	0.2	0.1
1	0.3	0.2	0.7	0.5
2	0.5	0.6	1.4	1.1
3	1.1	0.9	2.3	1.8
4	2.1	1.8	3.5	2.7
5	3.4	2.8	5.4	4.2
6	5.7	5.2	8.9	6.9
7	9.1	8.7	13.8	10.9
8	14.5	14.7	21.0	17.4
9	24.2	23.0	30.5	26.0
10	37.8	36.9	42.5	37.9
11	58.5	58.3	57.5	53.8
12	85.6	86.2	77.2	74.4
13	100.0	100.0	100.0	100.0
Nos. of Students	6613	5464	6234	5097

[a] Comparisons should only be made within each age level as different items were used for the 10-year-old and the 14-year-old students.

could have masked differences on particular tasks or on specific items within tasks. The individual link items were considered for the 14-year-olds to determine which tasks and items were responsible for the difference found between 1975 and 1980.

When the wide range of tasks shown in Table 9.10 is considered, it will be noted that differences between the proportions of 10-year-olds answering each of the link items correctly were small and certainly not significantly different for the two test administrations in 1975 and 1980. This indicated that the overall result that there were no differences in the performance of the 10-year-old students between 1975 and 1980 had not masked differences on a specific task or individual item basis. Once again, what was notable about the individual item results was their consistency between the two test administrations.

In the case of the 14-year-old students, the difference found for the link items sub-test as a whole was significant at the five per cent confidence

Table 9.10 Percentages of 10-Year-Old Students Responding Correctly to Link Items Between 1975 and 1980

Item No. in Test		Description of Task Tested	% Correct[a]	
1975	1980		1975	1980
3	2	Recognize spatial representation of fractions	87	88
4	3	Estimate in metres a length less than 10m	28	26
14	7	Count and order objects less than 10	94	96
8	8	Count the number of objects less than 20	98	99
29	9	Read time on a watch face with Arabic numerals	71	71
33	20	Add time using only one operation	69	71
19	21	Add two 2-digit numerals in solving problems	94	94
22	22	Add money in amounts less than 20¢	95	96
27	24	Add or multiply money less than $1	90	91
25	25	Add or multiply four 2-digit numerals in solving problems	89	89
30	26	Solve everyday problems requiring 2 operations	76	77
31	32	Solve everyday problems in metric units with 1 operation	69	69
9	35	Measure the length of an object in cm	96	94
Mean percentage of students correct			81	82

a None of the differences was significant.

level. Thus it might be anticipated that performance on at least some of the specific tasks and individual items would vary between 1975 and 1980. Table 9.11 presents the results for the link items and indicates the nature of the task tested in each case.

It will be noted from Table 9.11 that the proportions of 14-year-olds correctly answering these items were significantly higher in 1980 for five of the thirteen items. Differences on the remaining eight items were not significant, as approximately a five per cent difference was required for significance when item facilities were 70 to 80 per cent, and a six per cent difference was necessary with item facilities of 50 to 60 per cent (see Table 3.6). The items exhibiting significant differences were those concerned with reading maps and scales, with volume and with use of percentage. However, when all the link items were considered together, there was no significant difference in the mean proportions of correct responses which were 79 per cent in 1975 and 82 per cent in 1980.

Table 9.11 Percentages of 14-Year-Old Students Responding Correctly to Link Items Between 1975 and 1980

Item No. in Test		Description of Task Tested	% Correct[a]	
1975	1980		1975	1980
2	3	Add & subtract with time	83	85
1	4	Calculate equivalent areas	86	86
6	5	Read a map to obtain distance	79	84[a]
4	6	Add & subtract common fractions	63	58
14	12	Measure the length of an object in mm	90	92
8	21	Read time on a watch face without Arabic numerals	89	88
9	22	Read a bar graph and scale	88	90
29	24	Add & subtract with money amounts less than $1	89	91
20	26	Multiply or add with fractions to solve everyday problems	92	91
27	28	Read mass indicated on scales in gm	78	83[a]
31	33	Compare volumes of cuboids	57	63[a]
28	34	Use per cent to calculate simple interest	70	78[a]
33	35	Estimate approximate result of division involving more than 2 digits	68	75[a]
Mean percentage of students correct			79	82

a The differences between these proportions were significant in that they were greater than the sum of two standard errors for each of the 1975 and the 1980 test administrations.

Age Comparisons in Performance: 1980

A group of twelve items was selected to provide a direct basis of comparison of student performance in 1980 between the 10-year-old and the 14-year-old students. Eight of the twelve items were those already described as the basic items testing the four operations. As the items had to be used at both age levels, it was inevitable that some would tend to be selected from the more difficult items in the numeration test prepared for the 10-year-old students and the easier items in the test for the 14-year-olds. Table 9.12 indicates the nature of each of the items and the proportions of students at each age level who responded correctly to the item.

Three of the age anchor items involved the measurement of time. Whereas 88 and 95 per cent of the 14-year-olds were respectively able to write the

Table 9.12 Percentages of Students at Both Age Levels Responding Correctly to Items Common to Both Tests

Item No. in Test		Description of Task Tested	% Correct	
10 y.o.	14 y.o.		10 y.o.	14 y.o.
1	*1*	Recognize the value of a digit from its position in a numeral	91	97
10	*8*	Recall addition facts for 2 numerals each less than 10	97	99
11	*9*	Recall subtraction facts related to the addition of 2 numerals each less than 10	92	97
12	*10*	Recall multiplication facts for 2 numerals each less than 10	87	97
13	*11*	Recall division facts related to the multiplication of 2 numerals each less than 10	80	95
14	*17*	Read a calendar	74	90
16	*13*	Add two numerals each of 2 digits	88	94
17	*14*	Subtract two numerals each of 2 digits	74	92
18	*15*	Multiply a 2-digit numeral by a 1-digit numeral	60	85
19	*16*	Divide a 3-digit numeral by a 1-digit numeral to give a 2-digit integer	66	93
31	*25*	Calculate difference between times less than 1 hour shown on a circular watch-face and a digital clock	71	95
33	*21*	Read time from a watch-face without Arabic numerals	62	88
Mean of all 12 items			79	94

time shown on a watch-face and calculate the difference between two times displayed (one watch-face and one digital), only 62 per cent of the 10-year-olds could write the time and 71 per cent were able to calculate the time difference of five minutes. In interpreting a calendar, the proportions of 10- and 14-year-old students correct were respectively 74 and 90 per cent. Overall it was found that 15 per cent more of the 14-year-olds were able to answer correctly the anchor items, as can be observed in the mean proportions correct of 79 per cent for the 10-year-olds and 94 per cent for the 14-year-olds. This difference would probably have been greater if a ceiling effect caused by almost all the 14-year-olds being correct on some items had not existed.

Table 9.13 Cumulative Percentages of 10- and 14-Year-Old Students Correct on Age Anchor Items Sub-test

No. of Items Correct	Proportions of Students 10-Year-Old Cum. %	 14-Year-Old Cum. %
0	0.2	0.0
1	0.6	0.0
2	1.1	0.1
3	2.7	0.3
4	5.3	0.4
5	8.8	0.7
6	13.2	1.1
7	19.8	2.0
8	28.4	3.4
9	40.7	7.9
10	57.8	17.7
11	78.8	44.4
12	99.8	99.9
Number of Students	5466	5097
Missing Data (No of Students)	6	6
Percent Achieving Mastery	72	97

The twelve age anchor items were also considered as a sub-test and the proportions of students obtaining the possible range of scores from zero to twelve are shown in Table 9.13. The mastery level required for this twelve item sub-test was eight items correct, and this score was attained by 80 per cent and 98 per cent of the 10-year-old and the 14-year-old students respectively. Thus by this measure the 14-year-old students were superior by a margin of 18 per cent in coping effectively with the twelve age anchor items. This result was similar to that obtained on a different but similarly-intended sub-test of items common to the tests used at both age levels in the 1975 testing program. In that case 70 per cent of the 10-year-old students and 94 per cent of the 14-year-old students achieved mastery of the sub-test (Bourke and Keeves, 1977:77).

Summary

Students at both age levels exhibited a range of performance for the different tasks tested by means of sub-tests. In the case of the 10-year-olds, performance ranged from 67 per cent of students successful for the computation subtest on the ability dimension to 95 per cent successful on the money sub-test and 98 per cent for the sub-test involving addition. For the 14-year-old students, the range was from 80 per cent who were successful when working with fractions and decimals to 99 per cent for addition with the proportion of successful students exceeding 90 per cent for most sub-tests. Overall 78 per cent of the 10-year-olds and 86 per cent of the 14-year-olds mastered the total test.

At the 10-year-old level the female students were more successful than the males on some items and sub-tests, particularly those concerned with computation with numbers in a classroom context. On the majority of items and sub-tests however, the level of performance did not differ for the 10-year-old male and female students. At the 14-year-old level the performances of the male and female students were generally very similar with the males performing better on six items and the females better on two items. The males also had a higher proportion of successful students on the sub-test involving measurement using graphs, maps and scales. For all other sub-tests and the total test performance between the sexes did not differ significantly.

There were also some small differences in the performance of the 10-year-olds depending upon whether they attended schools in metropolitan or non-metropolitan areas. Where differences existed they favoured the metropolitan students. There were no differences in the performance of metropolitan and non-metropolitan students aged 14 years.

There were much larger differences at both age levels according to the year level of students. The level of performance of students in higher year levels was considerably higher for both the 10- and the 14-year-old student samples. However there was no difference in the results for the total samples of students when compared with the original samples without replacement of schools or individual students. Although the replacement students tended to have a higher level of performance for the 10-year-old student samples, the small numbers of students involved did not alter the overall results by even one percentage point. There was no difference in performance between the original and replacement students at the 14-year-old level.

It is clear that there had been a significant increase between 1975 and 1980 in the proportion of 10-year-old students who were able to perform basic recall and computation tasks with the four arithmetic operations. On these tasks the 14-year-old students generally performed at the same level in 1975 and 1980 except that for three items requiring multiplication and division, performance was significantly higher in 1980.

When a wider range including some of the more difficult items common to both the 1975 and the 1980 testing programs were compared, it was found that the level of performance of the 10-year-olds was almost identical on both occasions, whereas there was a small but significant increase in the level of performance of the 14-year-olds. The small increase resulted from larger and significant increases in performance from 1975 to 1980 for five items involved with reading maps and scales, with volume and with use of percentage.

The expected higher level of performance of the 14-year-old compared with the 10-year-old students on the same items in 1980 was similar to that found in 1975. In 1980, 18 per cent more of the older students mastered the sub-test formed from these twelve items. When individual items were considered, in every case the 14-year-old students had a higher level of performance than the 10-year-old students. For items concerned with place value, with addition, and with subtraction, the differences were small (generally less than 10 per cent), but for items involving multiplication, division and the measurement of time, the differences were much larger, in most cases of the order of 20 to 25 per cent.

CHAPTER TEN

SUMMARY AND CONCLUSIONS

The Australian Studies in Student Performance project involved the testing of a total of some 10 500 students aged 10 years and 14 years in schools throughout Australia during October 1980. The tests, which were criterion-referenced in nature and were developed specifically for the project, were designed to assess each student's abilities in reading, writing and numeration. The actual tasks tested were basic in that they were considered to be essential or at least important for further learning at school and for everyday life outside the school. The project was a response to community concern about the literacy and numeracy levels of students and had the major purpose of providing information in task-specific terms on the proportions of Australian students of these ages who were able to function effectively in reading, writing and numeration.

Test Development

The first stage in developing tests for use in this study was to determine the most important objectives to be assessed and consequently the tasks that students should be asked to perform. Using the objectives developed for the 1975 ASSP program as a starting point, objectives and tasks were set out for the 1980 program. The development of the reading, writing and numeration has been described and explained in some detail in Chapter Two. The resulting structure of each test is now summarized.

The Reading Tests

The first section of the reading tests at both age levels was concerned with word attack skills and consisted of an item which required knowledge of the alphabet. The second and third sections required the comprehension of simple and complex or compound sentences respectively. The former was concerned with assessing a very basic level of reading considered to be essential and the latter dealt with sentences that were structurally more involved but which were considered to be within the experience of 10- and 14-year-old students and important for their further education.

The fourth section consisted of a paragraph comprehension exercise where the passages chosen were of a type commonly found in school readers or textbooks. This section thus set out to assess school or classroom based reading as did section five which contained items assessing a student's ability to use enquiry or reference skills in reading an encyclopaedia and using the contents and index pages of books.

The sixth and final section of the reading tests moved from classroom based reading materials to community based reading. A simulated newspaper was used to assess a student's abilities to comprehend articles and to use reference skills such as skimming, scanning and using a timetable, a directory and an index in a community rather than a classroom context.

The Writing Tests

The writing tests were designed to test two distinct types of objectives. One type of objective related to the ability to respond to family, school or community based writing demands such as writing messages, instructions, reports and letters, and included the use of social conventions in tasks such as addressing an envelope. The items used to test the performance of these tasks were assessed against specific criteria and were scored on a correct/incorrect basis.

The other type of objective concerned the abilities to express personal feelings, ideas and experience and these abilities were assessed by means of tasks such as writing a story, writing about personal experience and writing attitudes and opinions. These items were assessed by global scoring on a normative five-point scale. Although the importance of objectives in the area of personal writing of this kind was unquestioned, it was clear that specific criteria for 'correct' answers should not, and indeed could not, be used in this case.

The Numeration Tests

The numeration tests developed for this study were not intended to cover the mathematics or even the arithmetic syllabuses of primary and secondary schools, but to assess student performance in those aspects related to either school or community requirements for numeracy. These requirements were seen to include the manipulation of whole numbers and fractions using the four arithmetic operations and the reading and use of measurements of quantity, length, mass, money and time. Various contexts were used to assess the abilities of knowledge, computation and application involved in seven content areas.

The areas were number and numeration concepts, properties of numbers and operations, arithmetic computation, measurement, probability and statistics, simple geometric concepts, and business and consumer mathematics.

Sub-test Scores

For both the reading and numeration tests, sub-tests of items assessing particular form or content, abilities and usages were developed and used as a basis for summarizing and reporting results in these areas. Success or failure on each sub-test was judged by calculating cutting scores based on the concept that a successful student should be able to answer correctly at least 80 per cent of all possible items which could have been used in that specific sub-test. As the actual group of items used in a sub-test was intended to be a representative selection of all possible items, measurement error would be present. The cutting score in each case was set at a level that minimized errors of classification of students according to the parameters of guessing and carelessness determined.

For writing, the individual criteria scored for each of the criterion-referenced tasks were assessed to determine which criteria were to be considered essential for the successful completion of the task. When writing a personal message, for example, the 14-year-old students were required to state both where they were going and when they would return to be considered to have successfully completed the task.

The Student Samples

The designed samples at each age level for the 1980 ASSP project consisted of 1000 students in each state and 500 students in each territory, making a total of 7000 students in each sample. The achieved samples actually consisted of 5473 students aged 10 years (78 per cent of the designed sample) and 5103 students aged 14 years (73 per cent of the designed sample). Although large numbers of students were involved in the study, the designed samples were less than three per cent of the total numbers of 10-year-old and 14-year-old students throughout Australia. This relatively light sampling approach was adopted to contain costs and to minimize disruption to schools. The achieved samples were a little more than two per cent of the student populations at each age level.

As losses from the samples were larger than might have been expected, sources of possible sample bias in the achieved samples were investigated.

A clear distortion in the achieved samples for most states and territories was that between government and non-government schools with the government school sub-samples having by far the greater losses. Given the differences in performance found in the 1975 ASSP program between students attending government and non-government schools, it was considered necessary to correct for this imbalance in the 1980 samples. Consequently when the samples were weighted to correct for the disproportionate sampling between states and territories, government and non-government schools were weighted separately to remove the sample bias towards non-government schools.

Other indications of sample bias were also sought. Although no imbalance was detected in geographic distribution within capital cities or within country areas, there was an imbalance between capital city and country areas in the 10-year-old student sample with too few city students. There was also a small imbalance between the sexes with both samples consisting of too many male students. Based on the results of the 1975 ASSP program, these small imbalances would be expected to depress slightly the level of performance of the overall samples in the 1980 ASSP program. Conversely there were differences in the year levels of students between the 1975 and the 1980 programs which would be expected to inflate the performance level in 1980 for both samples. However an increase in the reported use at home of languages other than English by the 14-year-old students would be expected to depress the 1980 performance level when compared with the 1975 program by almost the same amount as the possible inflation associated with year level differences. The nett effect of these differences on the 1980 results would be very small.

Sample losses were also investigated from a state/territory perspective. When weightings related to the state or territory populations of 10-year-old or 14-year-old students were applied, sample losses from the smaller states and from both territories assumed a lesser role in influencing the overall Australian results than was the case for losses from the larger states. Consequently although there were considerable sample losses for Tasmania, the Australian Capital Territory and (at the 14-year-old level only) the Northern Territory, these losses could have had only a marginal effect on the reported Australian results.

Losses from the Victorian and Queensland samples were potentially more serious. Apart from a few test items, it was calculated that the possible sample bias resulting from sample losses for these two states was negligible in most cases and in every case less than 0.5 per cent in the overall Australian results. However losses from the Victorian sample of 14-year-old students suggested that for a small number of numeration test items the Australian estimates of the proportions of successful students in 1980 could have been inflated by one per cent. It should be remembered that the bias discussed here was the maximum possible or potential sample bias. There is no way of determining whether the small degree of potential bias described was actually present in the results or whether the changes recorded, particularly for Victoria, were unrelated to sample bias. However the general trend of the results reported could not have been affected in any significant way by the potential bias. Those differences reported between the 1975 and the 1980 results were also large enough not to be invalidated by the sample losses in 1980.

Performance in Reading

The 1980 Results

The performance in reading by the 10-year-old students on the various sub-tests ranged from 70 per cent of successful students for comprehending a prose passage to 99 per cent for comprehending simple sentences. Less than 80 per cent of the students were successful on the sub-tests concerned with use of reference materials and the classroom and community usage sub-tests, and more than 80 per cent were successful when comprehending complex sentences and identifying the meaning of words in context. The total test was mastered by 71 per cent of the 10-year-old students.

The proportions of successful 14-year-old students ranged from 72 per cent for the comprehending continuous prose sub-test (which included the prose passage items and some of the newspaper items) to 100 per cent for comprehending simple sentences. For most other sub-tests more than 80 per cent of the students were successful. Exceptions were the sub-tests concerned with comprehending newspaper articles and community usage. The total test was mastered by 83 per cent of the 14-year-old students.

It would seem that at both age levels the simplest of reading tasks could be performed by all or almost all students, and every task tested could be done by more than two-thirds of the students. As expected more 14-year-old students were successful than 10-year-old students when attempting the same sub-tests with greater differences for the more difficult reading tasks.

When the results obtained were considered according to the sex of students, the female students had a significantly higher level of performance than the male students, particularly at the 10-year-old level. There were no significant differences in performance with respect to school location, that is, whether students were from metropolitan or non-metropolitan areas.

Comparison with the 1975 and Other Reading Results

There were three groups of items available for comparisons in performance between the 1975 and the 1980 reading test administrations at each age level. First there were several common or link items used on both occasions. It was found that, although there were some significant differences in performance on the two occasions for individual items, overall the results were very similar for the two administrations. When these items were formed into sub-tests it was calculated that 89 per cent of the 10-year-old students reached the required level of performance on both occasions, and in the case of the 14-year-old students, 93 and 94 percent were successful in 1975 and 1980 respectively.

Secondly there was a more indirect comparison available by using the sub-tests involving the comprehension of a prose passage. The relative difficulty levels of these sub-tests and the prose passages used in 1975 had been established during field-testing. It was found that the level of performance of the 10-year-olds on this sub-test did not differ between 1975 and 1980, but that in 1980 an average of approximately five per cent more of the 14-year-olds were correct on each item than had been the case in 1975.

Thirdly the newspaper items formed a useful basis of comparison of results between 1975 and 1980. Although only a few items used in 1980 were identical with those in the 1975 tests, many of the other items were similar in content and complexity. There was a marked increase in the level of performance of 10-year-old students between 1975 and 1980. Whereas only 27 per cent of the 10-year-olds had been successful on the newspaper items in 1975 compared with 53 per cent successful on the test as a whole, in 1980 the proportion of successful students was 71 per cent, the same as the proportion of successful students on the test as a whole. There was also an increase in the proportion of successful 14-year-old students on the newspaper items, although not as large as that reported for the 10-year-olds. In this case 64 per cent of the 14-year-olds were successful in 1975 compared with 75 per cent in 1980 on an equivalent newspaper reading sub-test.

One possible explanation for these large changes in student performance, especially at the 10-year-old level, is that considerably greater emphasis may have been given to newspaper reading and comprehension in schools throughout Australia since 1975.

One other comparison in reading results that should be mentioned concerns the prose passage used at the 14-year-old level. This passage had previously been used in 1970 by the IEA to assess student performance at age 14 years in England, the United States of America and New Zealand. In the 1975 ASSP study it was noted that the paragraph reading performance of Australian students in 1975 did not differ from that of students in England or the United States of America in 1970 but was lower than the performance of New Zealand students. The 1980 results indicated that the Australian students had a higher level of performance than the students in England had shown on the same passage in 1970, but did not differ statistically from the 1970 student performances in the United States of America or New Zealand. Clearly it would be of interest to know the performance level on this passage of present-day students in other English-speaking countries.

Performance in Writing

The 1980 Results

At the 10-year-old level, student performance on the criterion-referenced tasks ranged from 57 per cent for writing a letter for information to 93 per cent for writing a personal message within a family context. Whereas 87 per cent of the 14-year-old students were able to write a letter for information, their performance ranged from 54 per cent for writing directions to 97 per cent for writing a personal message. Thus at both age levels more than nine out of ten students could write a personal message but this was not the case for writing a telephone message. For this task more than nine out of ten 14-year-olds but fewer than seven out of ten 10-year-olds were successful.

For the globally scored tasks, the 10-year-old students wrote more when telling a story than when writing about themselves, and the 14-year-old students also wrote more for the story than for writing opinions which also had the highest spelling error rate. At this age level letter and story writing had approximately the same spelling error rate. The 10-year-olds had a higher spelling error rate when writing about themselves and writing a story and a lower rate when writing a report.

There was a consistent difference in writing performance between the sexes at the 10-year-old level which favoured the female students. The difference was less consistent at the 14-year-old level but the male students did less well than the females on the task of writing a letter requesting information and on all the globally-scored tasks. The only significant difference depending upon school location was for the task of writing a telephone message which was competently done by a greater proportion of the 10-year-old non-metropolitan students.

Comparison with the 1975 Writing Results

It was possible to make a few direct comparisons of student performance in writing in 1975 and 1980. For the task of writing a telephone message which was identical on both occasions, a slightly lower proportion of 10-year-olds were successful in 1980 than in 1975 and a slightly higher proportion of 14-year-olds. However neither of the differences was significant. The task of writing a personal message was also the same for the 10-year-olds in 1975 and 1980, and there was a significant improvement in performance with 86 per cent completing the task successfully in 1975 compared with 93 per cent in 1980. One other direct comparison could be made using the task requiring the writing of directions at the 14-year-old level. In this case there was no significant difference in performance between 1975 and 1980.

The 1975 and 1980 tasks of writing a letter for information were not identical but they included two common sub-tasks: these were writing a greeting at the beginning of the letter and giving a name at the end. Although the performance of the 14-year-old students did not differ on these two sub-tasks between 1975 and 1980, the performance of the 10-year-old students was higher on both sub-tasks in 1980. Whereas in 1975 only 63 per cent of the 10-year-olds wrote a greeting and 71 per cent gave their name, the corresponding sub-tasks were successfully completed in 1980 by 90 per cent and 80 per cent respectively.

Performance in Numeration

The 1980 Results

The proportion of successful 10-year-old students on the numeration sub-tests ranged from 67 per cent for computation to 98 per cent for addition. More than 80 per cent of these students were also successful on most of the 11 remaining sub-tests, the exceptions being the sub-tests concerned with general measurement, measurement of time and community usage where in each case the proportion of successful students was between 70 and 80 per cent. The total test was mastered by 78 per cent of the 10-year-olds.

In the case of the 14-year-olds, the proportion of successful students had a much smaller variation, ranging from 80 per cent for the fractions and decimals sub-test to 99 per cent for addition. Most of the other 12 sub-tests were handled successfully by at least 90 per cent of the 14-year-olds, the exceptions being the sub-tests concerned with general measurement, the ability of application and community usage. For these three sub-tests the proportion of successful students was between 80 and 90 per cent. The total test was mastered by 86 per cent of the 14-year-old students.

When the numeration test results of the male and female students were compared, it was found that, at the 10-year-old level, the female students generally outperformed the males. This was particularly so when computing with whole numbers in a classroom rather than in a community context. At the 14-year-old level the results for male and female students were generally closer but where larger differences did occur for individual items they more frequently favoured the males. The only sub-test where the male students had a higher level of performance than the female students was the general measurement sub-test which required the reading of graphs, maps and scales. There were no differences between the sexes on other sub-tests at the 14-year-old level.

There were also small differences in the performance of the 10-year-olds favouring students at metropolitan schools over non-metropolitan schools, but the differences did not carry over to the 14-year-old level. Larger differences in performance existed at both age levels according to the year level of students. More students in higher year levels were successful than students in lower year levels. There were also large differences in the proportions of successful 10-year-old and 14-year-old students on items used at both age levels. On average about 15 per cent more of the 14-year-olds were successful than 10-year-olds.

In general high proportions of students, especially those aged 14 years, were able to achieve success on the tasks assessed by the numeration test. Perhaps the only areas where this would seem not to be the case were those concerned with telling the time and working with time (the relatively lower proportion of 72 per cent of 10-year-old students was successful on these tasks), and, more particularly, the area of computation using the four arithmetic operations (67 per cent were successful). It is considered essential that all students in normal schooling be able to perform these tasks quickly and accurately before they leave the primary school.

Comparison with the 1975 Numeration Results

There were two bases of comparison of the 1975 and the 1980 ASSP results in numeration. First there were the eight basic items concerned with the four operations: four items involved only knowledge of tables, and four items each required the use of one of the four arithmetic operations, the operation required being specifically stated. The same items were used at both age levels. It was clear that there had been an improvement in performance at both age levels on these basic items between the 1975 and the 1980 administrations. The improvement was particularly marked at the 10-year-old level where there was a significant difference for six of the eight items. At the 14-year-old level where performance on these basic items in 1975 was already high, there was a significant improvement on three items in 1980, and in no case at either age level was performance lower in 1980 than it had been in 1975.

The second basis of comparison of the 1975 and the 1980 results consisted of thirteen items which were more representative of the numeration tests than the eight basic items just discussed. In this case different items were used at the 10-year-old and the 14-year-old levels. The performances of the 10-year-old students not only did not differ on these items in the two administrations, but were remarkably consistent when the score distributions were considered. This result was not repeated by the 14-year-old students who did perform at a significantly higher level in 1980 than in 1975 on a subtest formed from all thirteen common items. When individual items were considered, student performance in 1980 was significantly higher on five items with no significant differences for the other eight items. The items where a higher proportion of 14-year-olds were successful in 1980 dealt with reading maps and scales, calculating volume and using percentages.

Opportunities for Further Research

The data that have been collected and reported here contain a considerable amount of information on student performance on specific reading, writing and numeration tasks. There are opportunities to exploit these data in various ways to provide some insights into what might be done to assist students with learning problems. For example, responses to the writing tasks could be analyzed in detail structurally or from a different conceptual viewpoint to that used in this study. The individual errors and patterns of errors made in responding to the reading and numeration test items could also be analyzed in an attempt to identify general and specific problems that students had. The identification of these problems would make it possible to develop materials specifically designed to assist students to overcome such problems.

Final Comments

It is neither necessary nor desirable for the authors of this report to indicate whether they think that the performance levels described are satisfactory on unsatisfactory. What is suggested is that those persons interested in student performance in the areas of reading, writing and numeration should consider the relevance for literacy and numeracy of the tasks selected, the appropriateness of the items used to test student performance, and the results for the 10-year-old and 14-year-old students, before deciding for themselves their attitudes towards the results reported. Clearly while some will wish to emphasize the high proportions of students who could do most tasks, others will be concerned about the educational and social consequences for the students who were unable to do some tasks. We do not believe that everyone evaluating the results of the study will reach the same conclusions.

One aspect of the results obtained in this study was noteworthy regardless of whether the reported performance levels are considered to be adequate. There was no significant area in any of the assessments made where student performance at either age level had declined in the five years between 1975 and 1980. In fact there were a number of specific areas where student performance was higher in 1980 than it had been in 1975. In particular the better results in 1980 for newspaper reading by 10-year-olds, for comprehending prose by 14-year-olds, for writing a personal message by 10-year-olds, for the basic use of the four arithmetic operations by 10-year-olds, and for a range of other numeration tasks by 14-year-olds should be noted.

At the basis of the study reported here is the belief that every student in normal schooling in Australia should have every opportunity to become literate and numerate, and that deliberate actions are necessary to maximize each student's chance of success in this endeavour. The assessment of the degree of success achieved by these actions is seen as a necessary component of the educational process. To take this position is not to deny the importance of other aims and purposes of schooling associated with development of personality and individuality. In fact the abilities to read, write and calculate are considered to be prerequisites if individuals are to gain access to concepts, theories, and processes which in turn provide them with the means to structure their experiences in ways that develop understanding of themselves and of society.

REFERENCES

Australian Bureau of Statistics.
1976 1976 Census: Population and Dwellings in Local Government Areas and Urban Centres (Preliminary). Catalogue Nos. 2401.0 to 2408.0. Canberra: Commonwealth Government Printer.

Australian Bureau of Statistics.
1979 Schools Australia 1979. Catalogue No. 4202.0. Canberra: Commonwealth Government Printer.

Australian Council for Educational Research.
1976a Development of the Reading Tests. Melbourne: ACER.

Australian Council for Educational Research.
1976b Development of the Writing Tests. Melbourne: ACER.

Australian Council for Educational Research.
1976c Development of the Numeration Tests. Melbourne: ACER.

Blazely, L.D.
1978 1977 Survey of Reading Skills of 14-Year-Old Tasmanian Students. Research Study No. 34. Tasmania: Education Department.

Bourke, S.F.
1981 Development of the Numeration Tests. Melbourne: ACER.

Bourke, S.F. and J.P. Keeves.
1977 The Mastery of Literacy and Numeracy: Final Report. Australian Studies in School Performance Volume III. ERDC Report No. 13. Canberra: AGPS.

Bourke, S.F. and R. Lewis
1976 Literacy and Numeracy in Australian Schools: Item Report. Australian Studies in School Performance Volume II. ERDC Report No. 9. Canberra: AGPS.

Broadby, F.M.
1977 Survey of Basic Skills of Reading of 10-Year-Old Tasmanian Students. Research Study No. 27. Tasmania: Education Department.

Broadby, F.M.
1978 1977 Survey of Basic Number Skills of 10-Year-Old Tasmanian Students. Research Study No. 35. Tasmania: Education Department.

Cadman, A.G. (Chairman).
1976 Learning Difficulties in Children and Adults. Report of the House of Representatives Select Committee on Specific Learning Difficulties. Canberra: AGPS.

Calfee, R.C.
1978 'A proposal for practical (but good) research on reading.' Pp. 426-437. in L.J. Chapman and Pam Czerniewska (eds.), Reading from Process to Practice. London and Henley: Routledge and Kegan Paul.

Carroll, J.B.
1975 *The Teaching of French as a Foreign Language in Eight Countries*. Stockholm: Almqvist and Wiksell.

Cheyney, A.B.
1971 *Teaching Reading Skills Through the Newspaper*. Newark, Delaware: IRA Service Bulletin, Reading Aids Series.

Cronbach, L.J.
1964 *Essentials of Psychological Testing*. New York: Harper and Row.

Davidson, G.S. (Chairman)
1981 *Preparation for the Workforce*. Senate Standing Committee on Education and the Arts. Canberra: AGPS.

Dodson, H.L.
1976 *Reading Test for 14-Year-Old Students 1975 - Relationships Between High and Low Scores and Some Selected Factors*. Research Study No. 25. Tasmania: Education Department.

Glaser, R. and A.J. Nitko.
1971 'Measurement in Learning and Instruction.' Pp. 625-670 in R.J. Thorndike (Ed.) *Educational Measurement* (2nd ed.) Washington: American Council on Education.

Guilford, J.P. and B. Fruchter.
1978 *Fundamental Statistics in Psychology and Education*. (Sixth Edition). New York: McGraw-Hill Company.

Hays, W.L.
1963 *Statistics for Psychologists*. New York: Holt, Rinehart and Winston.

Hocking, H.
1976 *Survey of Basic Skills of Reading and Numeracy of Students Entering Technical and Further Education Institutions in Tasmania*. Research Report No. 24. Tasmania: Education Department.

Hunter-Grundin, E.
1979 *Literacy - A Systematic Start*. London: Harper and Row Ltd.

Izard, J.F.
1978 'Reliability of Criterion-Referenced Tests.' Collected papers of the Australian Association for Research in Education National Conference. Perth: AARE.

Keeves, J.P. and S.F. Bourke.
1976 *Literacy and Numeracy in Australian Schools: A First Report*. Australian Studies in School Performance Volume I. ERDC Report No. 8. Canberra: AGPS.

Kish, L.
1965 *Survey Sampling*. New York: John Wiley.

Kriewall, T.E.
1972 'Aspects and Applications of Criterion-Referenced Tests.' Paper presented to the Annual Meeting of the AERA. Chicago.

Marshall, J.L.
1975 'The Mean Split-half Coefficient of Agreement and its Relation to Other Test Indices: a Study Based on Simulated Data.' University of Wisconsin, Madison, Wisconsin.

Miller, G.A. (ed.).
1973 *Linguistic Communication: perspectives for research*. (Report of the Study Group on Linguistic Communication to the National Institute of Education). Newark, Delaware: International Reading Association.

Millman, J.
1973 'Passing Scores and Test Lengths for Domain-Referenced Measures.' *Review of Educational Research*, 43: 205-216.

Millman, J.
1974 'Criterion-Referenced Measurement.' Pp. 311-397 in W.J. Popham (Ed.). *Evaluation in Education: Current Applications*. Berkeley, California: McCutchan Publishing Co.

Mills, J.M. and F. Holzer.
1981 *Development of the Reading Tests*. Melbourne: ACER.

Morgan, G.
1979 *A Criterion-referenced Measurement Model with Corrections for Guessing and Carelessness*. Occasional Paper No. 13. Melbourne: ACER.

Myers, R. (Chairman).
1980 *Technological Change in Australia*. Volume One: Technological Change and Its Consequences. Canberra: AGPS.

National Assessment of Educational Progress.
1974 *Reading Objectives, second assessment*. Denver, Colorado: NAEP.

New South Wales Department of Education.
1976 *Enquiry into Literacy and Numeracy*. NSW: Department of Education.

Nitko, A.J.
1970 'Criterion-Referenced Testing in the Context of Instruction.' Paper presented to National Council on Measurement in Education Symposium 'Criterion-Referenced Measures: Pros and Cons.' New York.

Oppenheim, A.N.
1968 *Questionnaire Design and Attitude Measurement*. London: Heinemann.

Popham, W.J. and T.R. Husek.
1969 'Implications of Criterion-Referenced Measurement.' Journal of Educational Measurement 6, 1:1-9.

Renehan, W.R. and R.E. Wilkes.
1973 Report on Victorian Schools Assessment Programs 1969-1970. Melbourne: ACER.

Ross, K.N.
1976 Searching for Uncertainty. Melbourne: ACER.

Ryan, D.J.
1975 'The Nature of Reading.' Pp. 13-16 in The Literate Australian, papers and proceedings of a workshop, July 11-12, 1975. Melbourne: Council of Adult Education, Victoria.

Smith, F.
1978 Reading. Cambridge: Cambridge University Press.

Smith, Nita.
1978 'The use of the newspaper in the classroom.' Reading Education, 3, 2:45-62.

Spearritt, D.
1977 Measuring Reading Comprehension in the Upper Primary School. ERDC Report No. 11. Canberra: AGPS.

Stanyon, J.
1981 Development of the Writing Tests. Melbourne: ACER.

Survey Research Center.
1973 Osiris III: an integrated collection of computer programs for the management and analysis of social science data. The University of Michigan.

Teale, W.H. and R. Lewis.
1978 'Secondary school students' attitudes toward reading.' Paper presented at the 1978 National Conference of the Australian Association for Research in Education.

Thorndike, R.L.
1973 Reading Comprehension Education in Fifteen Countries. International Studies in Evaluation III. Stockholm: Almqvist and Wiksell.

Webster, A.H. (Chairman).
1980 National Assessment of Educational Progress. ERDC Report No. 26. Canberra: AGPS.

Williams, B.R. (Chairman).
1979a Report of the Committee of Inquiry into Education and Training. Volume I. Canberra: AGPS.

Williams, B.R. (Chairman).
1979b Report of the Committee of Inquiry into Education and Training. Volume 3. Canberra: AGPS.

APPENDIX ONE

MEMBERSHIP OF COMMITTEES

The Steering Committee

Chairman - Mr D.A. Swan, Director-General of Education,
Department of Education, NSW.

Mr M. Angus, Assistant Director (Research),
Department of Education, Western Australia.

Mr L.D. Blazely, Senior Superintendent of Research,
Department of Education, Tasmania.

Dr N. Bowman, Research and Evaluation Section,
Department of Education, Northern Territory.
(to 24 March 1980)

Dr R.K. Browne, AEC Secretariat.

Dr W. Donovan, Principal, Evaluation and Research Section,
ACT Schools Authority.

Mr S.S. Dunn, Chairman, ERDC.

Dr J.P. Keeves, Director, ACER.

Dr G.R. Maddocks, Assistant Director-General of Education,
(Curriculum and Planning), Department of Education,
Victoria (to 25 June 1980).

Dr W.M. Moore, Director, Division of Educational Research,
Department of Education, NSW.

Mr B. Morey, Chief, Education Planning Group, Federal
Department of Education.

Mr D. Pallant, Superintendent of Curriculum, Department of
Education, South Australia (from 25 June 1980).

Dr P. Tillett, Director, Research and Planning, Department of
Education, South Australia (to 25 June 1980).

Mr R. Warry, Director of Planning, Department of Education,
Queensland.

Mr Wee Cheng Sem, Senior Evaluation and Research Officer,
Professional Services Branch, Department of
Education, Northern Territory (from 24 March 1980).

Dr G. Whitehead, Assistant Director, Special Services Division,
Department of Education, Victoria
(from 25 June 1980)

The Technical Sub-Committee

Chairman - Mr L.D. Blazely, Senior Superintendent of Research,
Department of Education, Tasmania.

Dr W.F. Donovan, Principal, Evaluation and Research Section,
ACT Schools Authority.

Dr W.M. Moore, Director, Division of Educational Research,
Department of Education, NSW.

The Reading Test Development Committee

Chairman - Mr S.F. Bourke, Chief Research Officer, ACER.

Mr C. Browne, Senior Lecturer, English Department,
Burwood State College, Victoria.

Mrs D. Dungan, Head of the Junior School,
Lauriston Girls School, Armadale, Victoria.

Mrs F. Holzer, Research Officer, ACER.

Mr N. Matheson, Educational Consultant,
Catholic Education Officer, Victoria.

Mr J.M. Mills, Senior Research Officer, ACER.

The Writing Test Development Committee

Chairman - Mr M. Hamerston, Senior Lecturer,
Hawthorn State College, Victoria.

Dr E. Dines, Lecturer, University of Melbourne.

Mr P. Mildenhall, Department of Education, Victoria.

Mrs J. Mitchell, Co-ordinator, Special Education,
Methodist Ladies College, Victoria.

Ms J. Stanyon, Curriculum Officer,
Department of Education, Victoria

The Numeration Test Development Committee

Chairman - Dr J.F. Izard, Assistant Director, ACER.

Mr S.F. Bourke, Chief Research Officer, ACER.

Mrs F. Holzer, Research Officer, ACER.

Mr W. Newton, Curriculum and Research Branch,
Department of Education, Victoria.

Mr M. Quinn, Lecturer, Christ College,
SCV - ICE, Victoria.

APPENDIX TWO

LETTERS TO SCHOOLS AND PARENTS

The Australian Council for Educational Research Limited

EXECUTIVE
Emeritus Professor P H Karmel AC CBE BA PhD FACE FASSA (President)
A H Webster BA BSc FACE (Vice-President)
Professor D Spearritt MA MEd EdD (Vice-President)
S A Rayner MA MEd EdD FACE
G A Ramsey BSc DipT DipEd PhD MACE
DIRECTOR
J P Keeves BSc DipEd MEd PhD fil dr FACE FASSA

PO Box 210 Hawthorn
Victoria Australia 3122
Telephone (03) 818 1271
Cables Aceres Melbourne

1 July 1980

Dear Sir/Madam,

The Australian Council for Educational Research (ACER) has been commissioned by the Australian Education Council to undertake a study of aspects of school performance in reading, writing and numeration of 10- and 14-year-old students throughout Australia. It has been agreed that the ACER will approach schools directly to request their participation in the project.

Your school has been selected at random from the complete list of schools in your State as one of the schools to be requested to participate in the survey. May we therefore request your help and ask you to carry out a testing session with a group of 14-year-old students at your school? The testing session is planned for the week 6 to 10 October 1980 and will take no more than two hours to complete. It will involve a maximum of 25 students of the appropriate age at each school where there are sufficient numbers of students and fewer in smaller schools. We hope that the project will provide some factual evidence of use to those responsible for curricula by indicating the extent to which students aged 10 and 14 years can undertake ranges of tasks in the areas of reading, writing and numeration.

If you wish to participate in the project it is hoped that your students and staff will be interested by the rather different type of tests which have been especially prepared for this survey.

If they wish to receive them, schools participating in the project will be sent a copy of the results for their students as soon as possible after the tests are returned to the ACER. Subsequently in early 1981 a summary of the overall results of the project will also be sent to participating schools. Of course all test results and other information received will be treated as strictly confidential. No individual school or students results will be identifiable in any report produced as a result of the survey. I appreciate that some schools may be prepared to take part in the project but would have considerable difficulty in releasing a teacher for the two hours required to conduct the testing. In most cases we could arrange for assistance with testing to be provided at a time suitable to you. All costs of the project including postage will be met by the ACER.

2/.

A National Independent Educational Research Organization

I hope you will feel able to participate in this project. If you cannot, we will need to select a replacement school and this will reduce the accuracy of our carefully-drawn sample of schools. It would be helpful if you would nominate one of your staff, or yourself, as School Coordinator for the project and pass on the enclosed papers to him or her for information and, where appropriate, for completion and return. The Project Director, Mr Sid Bourke, would then communicate directly with the School Coordinator to arrange details for testing. If you have any questions about the project, please contact me or Mr Bourke at the ACER by letter or telephone (reverse charges). We look forward to your early reply to this request.

Yours sincerely,

J.P. Keeves

J.P. Keeves
Director

Enclosures:
1 Reply to this request
2 General Information
3 Memo to School Coordinator
4 Student Name Form
5 Reply-paid envelope.

The Australian Council for Educational Research Limited

EXECUTIVE
Emeritus Professor P H Karmel AC CBE BA PhD FACE FASSA (President)
A H Webster BA BSc FACE (Vice-President)
Professor D Spearritt MA MEd EdD (Vice-President)
S A Rayner MA MEd EdD FACE
G A Ramsey BSc DipT DipEd PhD MACE
DIRECTOR
J P Keeves BSc DipEd MEd PhD fil dr FACE FASSA

PO Box 210 Hawthorn
Victoria Australia 3122
Telephone (03) 818 1271
Cables Aceres Melbourne

1 October 1980

Name of child ____

Dear Parent,

This school is taking part in a national study of reading, writing and number work being conducted by the Australian Council for Educational Research (ACER). In the next several days your son or daughter may be one of those given tests, lasting less than two hours, in reading, writing and number work. We hope you will permit your child to take part in the study.

Your son or daughter will also be asked to write his or her country of birth and country of birth of each parent. Languages spoken at home will also be asked of each student. None of these questions is compulsory but if you do not mind him or her answering these questions it will help us in describing the overall sample of students we have. It would also help us if you would check that your child knows the answers to these questions.

All the information obtained from your child will be kept strictly confidential and no individual student information will be disclosed by the ACER. However if you have any objection to your child's inclusion in the testing program, please contact the school as soon as possible and he or she will not be tested.

If you have any questions about the study please contact your child's school or write direct to the ACER.

Yours sincerely,

JP Keeves

J.P. Keeves
Director

A National Independent Educational Research Organization

AUSTRALIAN STUDIES IN STUDENT PERFORMANCE

General Information

The Australian Studies in Student Performance is a project funded by the Australian Education Council (AEC).

The primary aim of the Australian Studies in Student Performance project is to gather information on achievement in aspects of literacy and numeracy of samples of 10 and 14 year old students throughout Australia and to make this information available to the public - and particularly to persons in the field of education.

Origins of the Project

In 1975 the House of Representatives Committee on Specific Learning Difficulties requested the Australian Council for Educational Research (ACER) to undertake a study of literacy and numeracy levels being achieved by Australian students. The results of this study were reported in three volumes in 1976 and 1977 and have been summarized in the Australian Education Review (No 11) published in 1978. The results reported the achievement of 10 and 14 year old students in Australia on a variety of tasks in the areas of reading, writing and numeration. However, it was not possible for this work to provide any information on changes in the levels of literacy and numeracy of the two age groups which might occur over time.

There is a continuing community interest in literacy and numeracy. The Australian Education Council, consisting of the Ministers of Education for all States and Territories and the Commonwealth, decided in 1979 that further information on the performance of Australian students was desirable. A Steering Committee comprised of representatives of each State and Territory Department of Education, the Commonwealth Department of Education and the Education Research and Development Committee was set up by the AEC to undertake the detailed supervision of the project to be conducted by the ACER. The assessment of students will begin in October 1980 and longer term plans are under consideration.

Purposes of the Program

The purposes of the survey program are summarized briefly as follows:

1. The program seeks to provide national data for some aspects of student performance in basic skills. Such information can be used to direct attention towards some areas which may be in need of improvement.

2. The existence of the program indicates that education authorities recognize the need for students to achieve at least a minimum level of literacy and numeracy.

Who Will be Tested

The samples of students to be tested will consist of 1,000 ten year old students and 1,000 fourteen year old students in each State and 500 at both age levels in each of the Northern Territory and the Australian Capital Territory. In each State 40 primary and 40 secondary schools will be chosen and 25 students randomly selected from all students at the appropriate age at the school will be tested where possible. In schools with less than 25 students at the appropriate age levels, samples will be smaller. Thus, up to 7,000 students at each of the two ages will be involved in the program. Such a sample will permit statements to be made about the performance of Australian 10 and 14 year old students.

The total numbers of students tested represent less than three per cent of the populations of 10 and 14 year old students in normal schooling in Australia. This light sampling approach is being used to provide, in an economical way, the evidence necessary to produce national estimates of student performance.

Schools will be invited to take part in the testing program. Those which decline will be replaced by a school of a similar type from the same geographic area.

What Will be Tested

In 1980, the three tests to be administered to the students at each age level will be in the areas of reading, writing and numeration and take no more than two hours of school time. The tests will be criterion-referenced so that every item will be related to a specific task. Because the tests will measure the performance of students against the task selected and not against each other, it is possible that all students could succeed on a specific task. The range of tasks to be included has been based on the advice of the AEC Steering Committee for the project, recent research into the amount of community agreement on these basic skills, deliberations of the Test Development Committees set up by the ACER, and the opinions of some teachers.

Reporting of Results

The major report of the study will indicate the proportions of all students able to perform the tasks tested and include a breakdown by certain demographic variables. This report should be available publicly in June 1981.

It is anticipated that a summary report of no more than ten pages will be prepared and sent to schools taking part in the project. This report will also be made available to others with a general interest in the study.

No individual school or student will be identifiable in any published report. Reports will be concerned with aggregates only.

APPENDIX THREE

ITEM REPORT

Proportions of successful students overall on The Reading and Numeration Tests (10R, 10N, 14R, 14N).

Table A3.1 Proportions of Students for Each Item of the Reading Test 10R[a]

Item	Name[b]	% Correct	% Incorrect	% No Attempt[c]
1	*Alphabet Completion	94	5	1
Simple Sentences				
2	*Birds on Roof	98	2	0
3	*Biggest Plane	98	2	0
4	*Children Swim	99	1	0
5	*Cutting Paper	99	1	0
6	*Fish in Bowl	97	2	0
Complex Sentences				
7	*Permission	89	11	0
8	*Gum Tree	91	9	0
9	*Lessons Japan	80	19	0
10	*Stayed Home	74	25	1
11	*School Team	91	9	1
Prose Comprehension				
12	Insect Attracted	90	10	0
13	Insect Light	79	21	1
14	Insect Disappears	86	14	1
15	Sundew Grow	75	24	1
16	Sundew Leaves	53	46	1
Reference Skills				
17	*Greenland	84	14	1
18	*Birds	59	36	5
19	*Fish Journey	80	18	2
20	*Air Travel	79	14	7
21	Famous Musician	84	9	7
Newspaper Reading				
22	*Plane Destination	55	44	1
23	Dolphin Location	88	9	4
24	Dolphins Love	86	12	2
25	Treasure Trove	81	16	3
26	Late Movie	92	4	4
27	*Pet Pups	82	11	7

* An asterisk indicates a task that was identical at both the 10-year-old and the 14-year-old age levels. (see Table A3.3).

a Proportions were calculated from the weighted numbers of students.

b Where percentages do not add up to 100, this has resulted from rounding error.

c In this column zero means less than 0.5 per cent.

Table A3.2 Proportions of Students for Each Item of the Numeration Test 10N[a]

Item	Name[b]	% Correct	% Incorrect	% No Attempt[c]
1	*Five Tens	91	8	1
2	Quarter Area	88	9	2
3	Height John's Dad	26	73	1
4	Road 10M Wide	82	17	1
5	Faye's 3/4 Circle	89	9	1
6	Maximum Pins	97	2	0
7	Running Fourth	96	4	0
8	No. Footballers	99	1	0
9	Tell Time - Numbers	71	26	3
10	*Addition/Tables	97	3	0
11	*Subtraction/Tables	92	8	0
12	*Multiplication/Tables	87	10	3
13	*Division/Tables	80	13	7
14	*Day 1st May	74	25	1
15	Total Money	93	7	0
16	*Addition/NAEP	88	12	1
17	*Subtraction/NAEP	74	25	2
18	*Multiplication/NAEP	60	31	9
19	*Division/NAEP	66	26	8
20	Time Arrive School	71	24	5
21	Total Balloons	94	5	1
22	Total Cost Choc.	96	4	0
23	Date in May	84	14	2
24	Total Cost Jelly Beans	91	9	1
25	Total No. Cards	89	10	1
26	Same No. Cakes	77	22	1
27	Joe's Cakes	91	8	1
28	Stamps for $2	88	11	2
29	Ice Cream Change	71	28	2
30	Points Red Team	78	21	1
31	*Time Difference	71	23	6
32	Medicine	69	25	6
33	*Tell Time - No Numbers	62	32	6
34	TV Program	91	6	4
35	Pencil Length	94	4	2

* An asterisk indicates a task that was identical at both the 10-year-old and the 14-year-old age levels. (see Table A3.4).

[a] Proportions were calculated from the weighted numbers of students.

[b] Where percentages do not add to 100, this has resulted from rounding error.

[c] In this column zero means less than 0.5 per cent.

Table A3.3 Proportions of Students for Each Item of the Reading Test 14R[a]

Item	Name[b]	% Correct	% Incorrect	% No Attempt[c]
1	*Alphabet Completion	98	2	0
Simple Sentences				
2	*Birds on Roof	99	1	0
3	*Biggest Plane	100	0	0
4	*Children Swim	100	0	0
5	*Cutting Paper	100	1	0
6	*Fish in Bowl	100	0	0
Complex Sentences				
7	*Permission	97	3	0
8	*Gum Tree	95	5	0
9	*Lessons Japan	92	8	0
10	*Stayed Home	94	5	0
11	*School Team	98	2	0
Prose Comprehension				
12	Use Parchment	96	4	0
13	Paper Discovered	87	13	0
14	Process Unsatisfactory	88	12	0
15	Paper Became	80	19	0
16	Main Point	43	57	0
17	Crude Means	78	22	1
Reference Skills				
18	*Greenland	95	5	0
19	*Birds	88	11	1
20	*Fish Journey	94	5	0
21	*Air Travel	96	4	1
22	Famous Writer	93	6	1
Newspaper Reading				
23	*Plane Destination	88	12	0
24	On Strike	74	25	0
25	Letters	64	35	1
26	Wave	84	16	1
27	Bumper	81	19	0
28	Golf Score	95	4	2
29	*Pet Pups	96	3	1
30	Perths Temp	87	13	1
31	Birds in Oil	86	10	4
32	Shot Birds	78	20	2
33	Tiger	69	27	4
34	Music Survey	71	24	5
35	Valour Award	88	5	7

* An asterisk indicates a task that was identical at both the 10-year-old and the 14-year-old age levels. (see Table A3.1).

a Proportions were calculated from the weighted numbers of students.

b Where percentages do not add to 100, this has resulted from rounding error.

c In this column zero means less than 0.5 per cent.

Table A3.4 Proportions of Students for Each Item of the Numeration Test 14N[a]

Item	Name[b]	% Correct	% Incorrect	% No Attempt[c]
1	*Five Tens	97	3	0
2	Six Tenths	81	29	0
3	Chemist Hours	85	16	0
4	Area Rectangle	86	14	1
5	Distance Dalton	84	16	0
6	Fraction 3/16	58	40	2
7	Bank Withdrawal	94	4	2
8	*Addition/Tables	99	1	0
9	*Subtraction/Tables	97	3	0
10	*Multiplication/Tables	97	2	0
11	*Division/Tables	95	4	1
12	Nail Length	92	8	0
13	*Addition/NAEP	94	6	0
14	*Subtraction/NAEP	92	8	0
15	*Multiplication/NAEP	85	13	2
16	*Division/NAEP	93	6	1
17	*Day 1st May	90	10	1
18	Distance Covered	77	20	3
19	Addition/Decimal	92	8	1
20	Subtraction/Decimal	88	12	1
21	*Tell Time - No Numbers	88	12	1
22	Apples Picked	90	10	0
23	Sue's Faster Time	67	31	1
24	Family Ticket	91	8	1
25	*Time Difference	95	4	1
26	Flour for Pies	91	9	1
27	Time TV Program	91	8	1
28	Read Scales	83	16	2
29	Record Change	90	9	1
30	Weight Difference	88	11	2
31	Time to Boston	84	12	4
32	Watch Change	88	11	1
33	Blocks in Box	63	35	2
34	Interest 10%	78	16	6
35	Division/Approximation	75	23	2

* An asterisk indicates a task that was identical at both the 10-year-old and the 14-year-old age levels. (see Table A3.2).

[a] Proportions were calculated from the weighted numbers of students.

[b] Where percentages do not add to 100, this has resulted from rounding error.

[c] In this column zero means less than 0.5 per cent.

APPENDIX FOUR

Included in this appendix are examples of students' writing which have been globally scored. Details of scoring methods are included in Chapter Two. The tasks set have been described in Chapter Eight.

The examples have been selected to show a range of responses for each task, from those responses receiving the minimum score of two to those receiving the maximum score of ten. Above each group of examples for a particular task are the scores attained and the percentages of students who received those scores or higher. For more open-ended tasks two examples in the middle grading (a score of 6) have been included, to give some indication of the variety of response. In the responses to the task which required a letter to be written, students' names, addresses and telephone numbers have been deleted to preserve the anonymity of the texts. The nature of the deletion has been recorded in capitalized type in the appropriate place. For instance a student's signature has been replaced by SIGNATURE.

Examples of Scripts of 10-Year-Old Students

Write About Yourself (Test 10W/A Task 4)

Example 1.	Score 10	3% of students attained this score
Example 2.	Score 8	21% of students attained this score or higher
Example 3.	Score 6	60% of students attained this score or higher
Example 4.	Score 6	60% of students attained this score or higher
Example 5.	Score 4	90% of students attained this score or higher
Example 6.	Score 2	All students attained this score or higher.

Example 1.

I enjoy a lot of things such as swimming high diving and gymnastics. My hobbies are collectings stamps, dolls, oriental things, smurfs, and wooden mice. My favourite sport is gymnastics but I think that I am better at ballet. I also love to perform in front of people for an example: now I'm in a dancing group for the Brisbane Estedford. I like to ice and roller skate although I'm not very good. At home I like to read and go on my trampoline, also whenever I get the chance I like to go swimming in my friends pool. But what I really want to do when I grow up is to be a zoologist or a biologist and study

Example 1 (continued)

nature. I like to borrow books from the library on animals and read all about them this way I get to learn a lot about animals and practise reading at the same time. There is one other thing I enjoy and that is going to a disco although I am to young for that now when I grow up I will be going to lots of discos.

Example 2.

I enjoy roller skating very much.

I go skating down the street.

I also go to a roller skating rink and hire some roller skates and I skate all around the rink.

I have my own skates but they are strap on's and you can't take them to a rink. because they only allow boot skates that is why I hire them.

I also enjoy netball. because it is a good game and I enjoy it a lot. I am in a team and I used to play every Saturday and have gala's some Sundays but I don't play any more because the season has finished.

But, I hope to play next season.

The End.

Example 3.

I engoy fishing, climbing, and making modles. I like modles because its lots of fun and its difficult. Fishing I like that for food and the fun. Climbings fun because sometimes I think I am going to fall and I feel glad when I make it. But it's expensive if you want the proper gear. Modling and fishing are expensive to but I still like them all for fun. I dont like the expence for boots parkers modles, glue and fishingrods.

Example 4.

I like going swimming on hot school days. After school I run home and put on my bathers and run down to the i c i wharfs where a lot of people swim. The water is very cool and I play chasy with my friends. When it gets dark I go home with my friends I haven't far to walk because I only live around the corner when I get home I have a hot showerer and then watch television.

Example 5.

I enyoy doing lots of things like stamp collecting, skateboarding, rollerskating, playing with my dog and friends but most of all I like swimming, riding my bike, playing hockey, cricket teeball and soccer.

Example 6.

I mail like to go swimming but it is not my favit hode. I like all difrent hodes I can name some but there is to meny to name. I can name two balla and swimming I ges but sometimes all the hodes ar broring Some ar good and some ar booring. but my hodes ar not that good.

Write a Report (Test 10W/A Task 3)

Example 1.	Score 10	5% of students attained this score
Example 2.	Score 8	23% of students attained this score or higher
Example 3.	Score 6	63% of students attained this score or higher
Example 4.	Score 4	93% of students attained this score or higher
Example 5.	Score 2	All students attained this score or higher.

Example 1.

The bean cycle

~~How beans~~ come ~~are to your table~~

One day the farmer decided to plant some bean seeds. a few months later the plant bore fruit. The farmer's wife noticed the tall bean plants and went out to get the beans. She decided to have them for dinner. Soon she had prepared them and cooked them in a pot. When the farmer tasted the beans he said it was delicous. ~~So he~~ The family agreed with him.

Example 2.

The Beans and Fred

Fred went out to the garden to plant some beans He planted about five bean trees and waterd them with fresh water. The next day they grew about 10 centermetres and Fred was very proud of himself. Then his wife Betty went out to pick some beans from the bean trees for dinner she cut them up nicely and washed them then put them in a pot to cook for 30 mins. Then the 4 children Fred and Betty sat at the table to eat

Example 3.

On Saturday father planted some
seeds in the garden. They grew
and grew until it was time
for them to be picked. So mother
picked the beens and cooked
them for tea that night.

Example 4.

MaKE your own tea

The man has put some seeds
in and the ~~po~~ plants are growing.
Now The Lady is taking The fruit
off the tree and puts the in the
cooking pot. And now She is cooking
them. ~~a~~ ~~And~~ now they are having
tea

Example 5.

The man is putting same bens in the grand
That day thay are big some the Mum
come and ges them and gose in sind
and cuts them up and cood them up
and had them for tea.

Write a Story (Test 10W/B Task 3)

Example 1.	Score 10	5% of students attained this score
Example 2.	Score 8	21% of students attained this score or higher
Example 3.	Score 6	60% of students attained this score or higher
Example 4.	Score 6	60% of students attained this score or higher
Example 5.	Score 4	90% of students attained this score or higher
Example 6.	Score 2	All students attained this score or higher.

Example 1.

One day I was playing soccer when I stumbled
over something, I picked it up it looked like
some kid has lost his toy flying saucer. I looked
at More closely and found there was a little
blot a light greenish colour. It turned around

Example 1 (continued)

It pulled out a thing like a twig and said "Open the door." It startled me that it was talking and so I opened the door for it.

It spoke again "My name is Richi I am from the place Zilror take me to your home"

I took Richi home and fired at least 20 questions at him. He only heard about 5 because he was swinging on the iron cord. His weight was to much for it. The iron fell. It landed on the cat's tail The cat leaped on the chandilier. It fell on the china crockery set. The cat leaped on our dog.

The dog started chasing the cat and made a mess of the kitchen Meanwhile Richi was having a fine time playing around the vase it crashed the water went everywhere It short circuited the T V set which caught on fire which caught on to the drapes. And there I was trapped by the fire. "Wake up Joe Wake up" Dad was telling me then It was only a dream.

Example 2.

Riki

I was walking along the road when suddenly I heard a whirring sound then this strange little creature came out and said "Hello I'm Riki, please do not be frightened.

So I took Ricki home. I was just walking up the stairs when Riki disappeared. "Come Up for supper Justen," called mum, So I went up stairs.

As I sat down crash Riki had just knocked the milkbottle, the bottle of champagn and he stood in the cherry pie. Almost instantly Riki disappeared again. "What on earth is that!" exclaimed dad. So I told dad all about Riki. After the discussion dad said "I'm afraid we cannot keep Riki. He will be too much of a bother." Heartbroken I went to my room, and I never saw Riki again.

Example 3.

One day a space creature came to our house. His name was Riki. He was very friendly. We asked him to have some tea with us. We sat down to the table. Before we could even touch the food Riki starting eating. We were very mad so we tried to catch him. But Suddenly he disappeared. We couldn't see him. By that time the house was a mess. Then I saw him and caught him. We put him in his space ship a pushed him back onto his home planet. We didn't see him again but we were still glad that he was gone.

Example 4.

My Friend, Riki

I have a friend and his name is Riki. Riki is nice but sometimes he plays tricks some of ~~x~~ the tricks are practical jokes and others are nasty tricks.

This is what happened when Riki visited me.

One day when I was walking home from school Riki to me and asked, "May I come and play at your house," and I answered, "Yes."

When we were there Riki started playing tricks. He swing on the light shade, broke a window, spilt everything on the table, knocked over a chair, broke the lamp shad and rolled up the carpet and the never appeared for his punishment.

This is my advice ~~never~~ never invite Riki to your house.

Example 5.

One day Riki came to see me while
Iwas put rocks arand my plant.
Riki braught rock for breat they weren't
rocks they were dimons and every
one was digind in my garded
for dimons they thats where I got
the dimons two ralibers came
to get the dimons so Riki took
them to the planit where he got
the dimons and left themthere.
Riki came back again and I said
what. Riki said Ithink earth people
call it gold I fellto the graundand
fell in a puddle Riki laughted
I was wet from head to toes
I went inside and had a shour
and put clean cloths on.

Example 6.

When Riki came a me
home he made a lot
of trombel. We ket hem
out of the hause and
he whent away. for good.
I was outsida and I saw
Riki agian but he wheant
home.

Examples of Scripts of 14-Year-Old Students

Write a Story (Test 14W/A Task 3)

Example 1.	Score 10	3% of students attained this score
Example 2.	Score 8	22% of students attained this score or higher
Example 3.	Score 6	56% of students attained this score or higher
Example 4.	Score 6	56% of students attained this score or higher
Example 5.	Score 4	87% of students attained this score or higher
Example 6.	Score 2	All students attained this score or higher.

Example 1.

Quietly Sue waited for Daniel and Mark to reach her. All three of them were clad in black clothing. As they stood in front of Sue's house, they quickly checked over their gear to make sure that nothing was missing. Hammer, rope, bird cages, halter and rubber band were all there. Ready now, they set off towards the deserted fish and chip shop which was about two blocks away.

On the way there, they talked in low whispers about the happenings of the last week. Everything had happened slowly, mysteriously. One by one, the neighbourhood pets had started disappearing. First Sue's talking Cockatoo had gone missing, then Daniel's welsh pony had been taken from its paddock. As well

Example 1 (continued)

as these pets, Mark's twelve racing pigeons had been taken from their perch and another neighbourhood pony had also gone missing. There was only one clue. At each house, a white truck had been seen. It was large enough to have all the pets concealed within it, so now Sue, Mark and Daniel headed for it, hoping to get their pets back.

Upon reaching the truck, they stopped abruptly as they heard a soft whinny. Daniel raced around to the back of the truck, hoping to be able to open the doors. Unfortunately they were locked, so taking the hammer from Sue, he gave the window a sharp hit, breaking the glass. Then he reached inside and opened the doors.

They were met by a stocky pony, who whinnied loudly as Daniel approached.

Silently they put halters on the ponies and put the pigeons in their cages. A rubberband over the cockie's beak stopped him talking.

Then triumphantly they led or carried their happy pets home.

Example 2.

"I have an idea," I said to my two best friends, Marina and Lotte.

"What is it?" they both asked anxiously. "Is it about all the pets that have been disappearing lately?"

"Yes. Along with the twelve racing pigeons and two ponies, a talking cockatoo has also gone missing," I answered them.

As they both listened attentively I told them about my suspicions of the white truck parked outside the empty shop. They both agreed with my suspicions, because every time an animal disappeared, the truck was sure enough parked there and then zoomed off after the crime.

We had an idea. We were planning to trick the ~~thiefs~~ thieves and make them fall into our trap. What we were thinking of was to buy a performing dog-something any animal thief would want to get his hands on.

You might say we used the dog as bait, but we didn't let him out of our sight for one minute. We walked around the neighbour-hood, bragging about Marina's new performing dog. We especially made sure that we stood outside the empty shop, where the white truck usually stands and bragged tremendously there, and told everyone every single detail needed to know.

That night we decided to wait for the thaives to come and steal the dog from us and sure enough at about ~~15~~ ~~f~~ fifteen minutes past midnight, ~~3~~ three men crept into the yard. ~~Als~~ ~~Kn~~ Knowing that we couldn't tackle them all

Example 2 (continued)

by ourselves we had called the police previously to be on guard around the front of the house.

Lotte crept around the front and told the police man about the thieves on the attack

Meanwhile the three men feeling proud of themselves and not knowing that any-one had seen them, stalked out through the front gate.

" Hold it there," yelled one of the policemen.

'Quick! Run for it, boys.' ordered the leader of the thieves They started to run for their lives, but not fast enough. The policemen had caught up to them after quite a chase and had hand-cuffed them.

Marina, Lotte and myself were feeling very proud of ourselves, and what better way to end the night by ~~even gett~~ receiving a small reward

Example 3.

Theres something very supisious about that white truck parked out side that empty shop' said Jane with a slight grin on her face. I agreed with her and so did Julie. We all hid in the bushes ~~at~~ across the road from the shop so we could try and have a look what was going on. Suddenly a man came out of the shop he opened the groll-a-dor and got in the white Truck and slowly back it through the opening and then the doors where closed again. We all turned to face each other.

Example 3 (continued)

I then said "now lets try to put all the pieces together on this mystery". Right said my two other friends "now in the past week" I said, "twelve racing pigeons, two ponies a talking cockatoo and other animals have gone missing. then we see a white truck parked out side a empty shop.

"If you askme" said Julie, "I think its all tied together somehow. I agreed with her. Then Jane said "come on and lets take a look. We ran across the road and looked in the shop window we could not see anything, then I notice that the door was slightly ajar. We all crept in had a look inside the warehouse. Inside were the ponies, pigeons, and the other animals that had been stolen. I told Jane to quickly run to the police station. He had just went out when one of the guys came in and saw us we tryed to run but they we holding a gun. They took us in to the warehouse and tied us up. Then the police came. There were cars, trucks and motorcycles everywhere. The arrested the robbers and we got a reward for saving the animals. If the police did not find the animals they were going to be sent overseas and sold for thousands of dollars each.

Example 4.

In my area too many pets had been disappearing lately, icluding my cockatoo who's name was Freddy. I was very angry so I called over two friends of mine, Tom and John.

We put all the clues that we had together and we found out that every time that a pet or two had disappeared, there was a white truck outside the empty shop down the corner. We got very angry so we decided to go and investigate.

When we got there it was about 4 o'clock in the afternoon. As we got, we saw the white truck coming and we hid behind a car. The truck stopped outside the shop, again, and a man came out. He unlocked the door of the shop and also of the truck. Then we saw him carrying a cage inside with my bird in it. I was very angry when I saw the way he carried my Fred.

I could not tell this to the police because they would not believe me.

Then Tom had an idea. He and John would sneak in at night time and pretend to be ~~thief~~ thieves. Then I would ~~go~~ call ~~to~~ the police and tell them what is going.

The police got there and surrounded the place. They crept in slowly and were very surprised

Example 4 (continued)

when they saw all the animals. The animals ~~who~~ were all in portable cages except the horses, and they all had price tags on them.

My bird was to be sold for a mere $12 while he was worth about $40. But, ofcourse! It was black market!

The police decided to stay there till the morning to catch the thieves and told us not to say anything to anybody, and so we did.

As the white truck came in the morning police surrounded them from every where. The whole gan was ~~[illegible]~~ caught thanks to me and my friends. My bird was also given back to me.

The police told us that these people had been pestering them for 3 months now and because we got all the animals back, there would be a reward for me and my friends. I wish something like this never happens again though, because it was all very dangerous.

* * *

Example 5.

One day my two friends and I went looking for all the missing pets that are missing and John my friend had an suspicion about a white truck out-side an empty shop, so we thought we would take a look. First we had a look in the truck and there was all of these cages in there but they were empty so we went around the back of the shop to see if we could find any thing but we didn't find a trace of an animal so we tried to get inside the shop and we did through a window that was unlocked.

We looked around for a bit and we couldn't find any thing until Mike my other friend found a door in the floor under a carpet, so we opened it up slowly incase somebody was down there but we opened it up and all we could see is cages so we went down there and we saw all the animals locked in these cages but they had padlocks on them and we couldn't get them off so we

Example 5 (continued)

found some scissors and tryed to break the wire and after a while it broke and we got all of the animals out and then we opened the door in the floor again and there was men up there so we couldn't get out but then Mike found a window in there and so John climbed out because he was the smallest and got the police and then the police came and got the animal thieves and let us out and got all the animals back to there owners and my two friends and I we were proud of from every body, they said we were brave little boys.

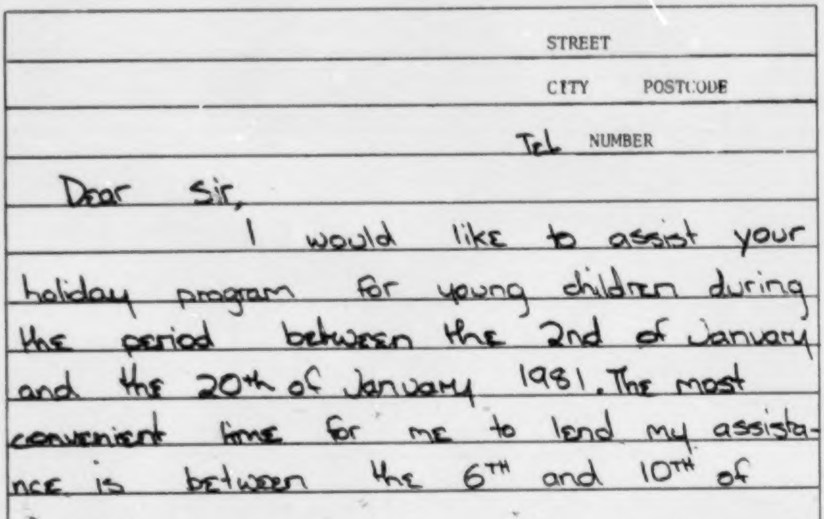
Example 6.

One day slit week we took to my friends and he said is twoponies went miss and our tourgh friens pigeons were sow we biaten that we would trid to find were they W ent to and we sow white truck parked out side the empty shop sow we were going to have and look what there were boing sow we ran. around to the back of the old shop went. we went we sow man with and gun and we hode to have and close look sow one of we inclimed and tree and made the tree becouse we fot and sat up the thee so we could get the man so we could get. but the onthe man could us and he took wis slide but one of our friend got a way. And he went to the tell the poiler but there would not belive them so the boy went to tell his mother but she bid not belive him sow he went back to the poiler and there went to have a look.

Write a Letter (Test 14W/A Task 2)

Example 1.	Score 10	4% of students attained this score
Example 2.	Score 8	23% of students attained this score or higher
Example 3.	Score 6	58% of students attained this score or higher
Example 4.	Score 4	84% of students attained this score or higher
Example 5.	Score 2	All students attained this score or higher.

Example 1.



STREET

CITY POSTCODE

Tel NUMBER

Dear Sir,

I would like to assist your holiday program for young children during the period between the 2nd of January and the 20th of January 1981. The most convenient time for me to lend my assistance is between the 6TH and 10TH of January.

My interests are quite varied the main one being animals and animal breeding. However I have had experience with children. I babysit quite often and have two younger brothers of age 2 and 6 yrs.

I have had organizing experience being in the Girl Guide movement for quite some time.

Yours Faithfully

SIGNATURE.

Example 2.

NAME
STREET
Bondara 2360
N.S.W.
19.10.80.

Dear Sir/Madam,

I have read your advertisement, advertising that you need help from teenagers in this area to assist you

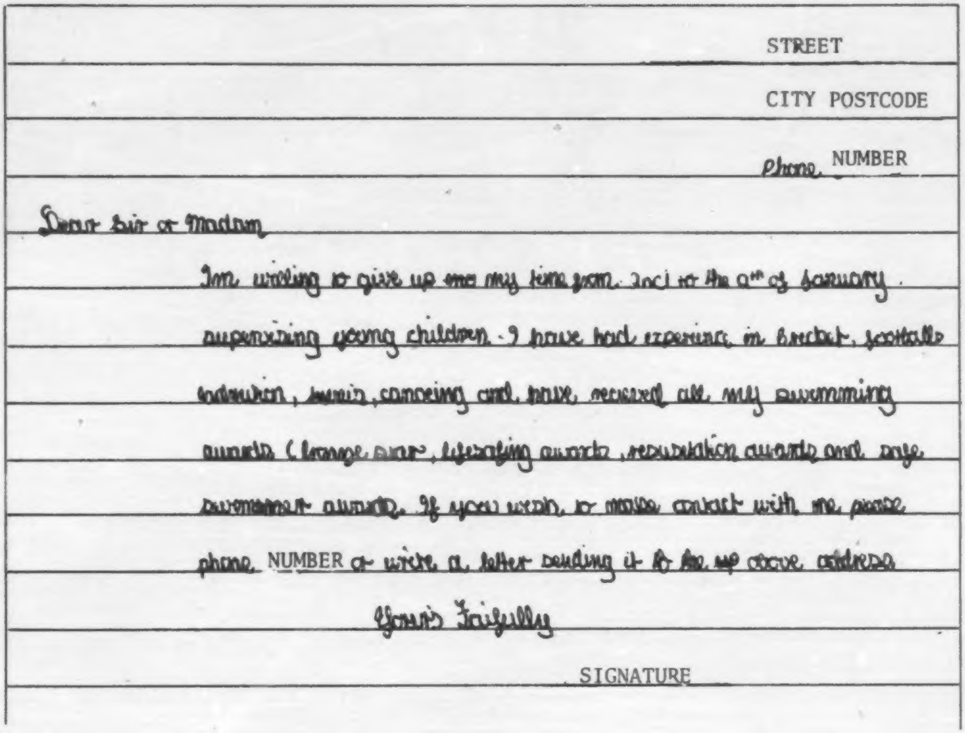
I will be able to assist you from 10·00 am till 4·00 pm from the 2nd January to the 10th.

My Interests are very few, but I may have some that will help. I like doing writing: typing, commercial art & also outdoor activities. I also like helping children do work, & also play. I have an interest in small children, I babysit whenever I can, & have a large family. I will help with anything I can for you

Yours Sincerely

SIGNATURE

Example 3.



STREET

CITY POSTCODE

Phone NUMBER

Dear Sir or Madam

Im willing to give up ~~ma~~ my time from. 2nd to the 9th of January. supervising young children. I have had experince in cricket, footballs badminton, tennis, canoeing and have recieved all my swimming awards (bronze star, lifesaving awards, resusitation awards and safe swimmer awards. If you wish to make contact with me please phone NUMBER or write a letter sending it to the ~~up~~ above address

Your's Faifully

SIGNATURE

Example 4.

181 Wombat Way
Bendora 5906.
TELEPHONE NO.
6·10·80.

Dear Sir/Madam.

I am answering your request for teenagers help in your holiday program. I get along well with children and have a good patience. I will be able to assist you from the following dates.

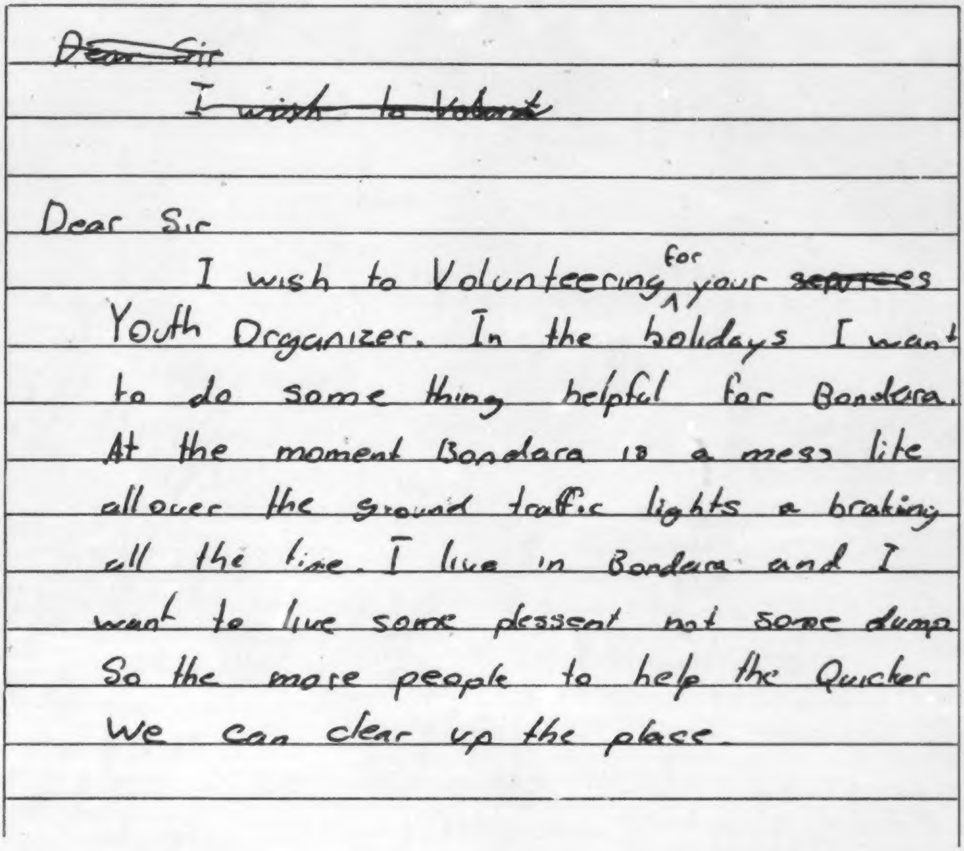
2nd - 9th. January
11th - 14th January.

If you need my services please reply and i will be willing to assist.

Yours faithfully

SIGNATURE

Example 5.



~~Dear Sir~~

~~I wish to Volunt~~

Dear Sir

I wish to Volunteering for your ~~services~~ Youth Organizer. In the holidays I want to do some thing helpful for Bondara. At the moment Bondara is a mess lite allover the ground traffic lights a braking all the time. I live in Bondara and I want to live some plessent not some dump So the more people to help the Quicker We can clear up the place.

Write Your Opinion (Test 14W/B Task 4)

Example 1.	Score 10	3% of students attained this score
Example 2.	Score 8	17% of students attained this score or higher
Example 3.	Score 6	49% of students attained this score or higher
Example 4.	Score 6	49% of students attained this score or higher
Example 5.	Score 4	80% of students attained this score or higher
Example 6.	Score 2	All students attained this score or higher.

Example 1.

The statement "Teenagers Have too Much Freedom" is not entirely correct. Some teenagers are allowed to do what they like but others are under the strict supervision of their parents and even their teachers.

Many teenagers are allowed to do what they like, when they like. This seems to occur in many families where both parents are working and there is no one at home to watch over the children. This may not be the parents' fault. They may both have low paying jobs, thus needing to work constently to make a living. It also may be a one-parent family where the sole parent must work and does not have the time to set down rules for the teenagers who are left at home. In some cases it may be impossible to place restricting rules on teenagers. In these cases it is up to the children themselves to determine what they should and shouldn't do and if they have the wrong idea they can get into trouble.

On the other hand, some teenagers do not get enough freedom. Some are kept at home and not allowed to pursue sporting interests or hobbies or even to visit their friends.

Example 1 (continued)

This is sometimes the parents' fault because they have the wrong ideas about discipline. But they are only trying to stop their children from being hurt or getting into trouble.

If more parents could find the median between absolute freedom and its opposite, imprisonment, then there would be less teenager's getting into trouble and the world would be a happier place.

Example 2.

When we say teenagers what do we mean? Is it boys and girls in senior Grades? Or is it people from 13-19? Either of the two they ~~does~~ not have as much freedom as they could or parents think. This child "Expert" says we have too much freedom, compared to what. To when they were kids. And haven't we all heard - "When I was your age ~~they~~ - I couldn't do this, but had to do that." Can't they see times have changed. Our world is rapidly changing around us. There have never been so many technological advances which change things. With new computers, school leaving teenagers ~~with~~ can't get a job so they have far more leisure time. Or the jobs are part time

Example 2 (continued)

so that they have increased leisure time. But it isn't just teenagers who have this freedom but working people too. Then there is the child in secondary school, they do not get all this freedom, during the week many families make their children do all their homework and some study on weekends, leaving maybe Saturday nights to go out. Even then not everyone goes out So come Saturday night. Not like "then" and every Saturday they went to the pictures, these days it costs far more Sunday's are never much, and the only freedom around is maybe TV watching, going for a ride, a game of sport or one of the machines at the local Pizza bar. That is not real freedom we are limited in just about all we do, ~~&~~ when money runs out, or we get masses of homework

We have no more freedom than ever. Not many parents I know have "lost control over theirs kids". It is just that the different forms of leisure and freedom have changed. The parents usually have say over to whose place ~~so~~ a child will go and

Example 2 (continued)

for how long. They supply money and will say when it stops and then you must work to get it (~~half the~~ wage less than half the adult wage)

It depends what you class as freedom!

Example 3.

To the Editor,

I saw the headlines in yesterdays News (8-10-80) about teenagers having toomuch. I am sick ~~and ti~~ of people alway accussing teenagers of being to blame. It is always the teenagers who get blamed for things that most often are not true. Teenagers are easy to blame because they are going through a hard time, they are turning from children to adults. I am sure that Dr. M Smith acted just the same as most youths do today when he was teenager, and I am sure that similar child "experts" said the same thing about teenagers. It is recorded ~~said much~~ that a Greek said much the same thing and that was 1,000 B.C.. You cannot blame the problems of today on teenagers. It could quite easily be said that teenagers have to little freedom. Most of the vandals and other hooligans today do not have too much freedom, they have to little. The parents have too much control over them. The freedom

Example 3 (continued)

they do have they miss use. These are the teenagers who get in trouble. I say teenagers ~~set~~ should have more freedom. This would develop a some responsiblity in them. Trust is what they need, if they have trusted they will act normally. So I say teenagers have to much freedom, they have too little. The people who write ~~that~~ that sort of thing should took at what they did as teenagers. They had too little freedom so everyone does.

Yours gratefully
, an angry teenager.

Example 4.

I don't really think that most parent have lost control over their children, at least not migrants anyway. Most greek, italian, etc. ~~children~~ teenagers aren't really allowed to go out alot in Australia because their parents are afraid that they might get mixed up with drugs and wrong people. For these reason many parents dont let their kids to go out by themselves. And since ~~kids~~ these teenagers don't go out by themselves then they stick with their parents and don't know what it's like to answer back and be rude to them.

But then again there are teenagers who are ~~alow~~ allowed to go out and they get mixed up with the wrong people and I think

Example 4 (continued)

that it's these sought of teenagers that have bad parents loose control over them.

I think that Dr. M. Smith is right in some way but then wrong again becouse not all teenagers have the same amount of freedom.

Example 5.

Know I dont think teenagers have too much freedom because they are just like normal people. They go to places with their friends and do other things.

But on the other hand some teenagers have too much freedom because they are allowed to go out as late at they want drinking and smoking and when they come home, they come home drunk and their parents don't really care and they don't get into trouble so they can do it again and again.

Example 6.

I think teenagers have just about all the freedom thay need. Thay are allowed to go to danres and pitchers just about as much as thay want. So I ~~beli~~ belive teenagers have just about all the freedom the want. People say thay have two much but thay are not teenagers. If thay were teenagers thay would say thay would not have enough freedom.

APPENDIX FIVE

SCORE DISTRIBUTIONS AND RELIABILITIES

READING AND NUMERATION

Test and Sub-Test Distributions
Reliabilities

WRITING

Reliabilities of Global Scores

Table A5.1 Distribution of Scores on Reading Sub-Tests and Total Test for 10-Year-Old Australian Students

No. of Correct Items	Percent of Students Weighted[a]											
	Ability: Comprehending Simple Sentences	Ability: Comprehending Complex Sentences	Ability: Prose Comprehension	Ability: Reference Skills (Classroom)	Form: Words in Context	Form: Reference Materials	Form: Continuous Prose	Usage: Classroom	Usage: Community	Link Items 1975/80	Age Anchor Items	Total Test
0	0.1	0.6	2.2	1.6	0.0	0.1	0.7	0.0	1.5	0.1	0.0	0.0
1	0.1	2.2	5.5	5.0	0.0	0.9	1.9	0.0	2.8	0.2	0.0	0.0
2	0.2	5.5	8.3	8.8	0.1	1.7	3.2	0.0	4.0	1.3	0.1	0.0
3	0.8	10.9	14.3	15.5	0.1	2.7	6.2	0.1	6.7	3.0	0.1	0.0
4	5.3	24.7	30.5	27.6	0.3	3.3	7.3	0.1	13.7	6.1	0.0	0.0
5	93.1	55.9	38.9	41.2	0.8	6.1	10.9	0.1	29.2	17.1	0.2	0.1
6					2.7	8.0	19.0	0.1	41.7	72.0	0.2	0.1
7					5.3	14.6	26.0	0.2			0.6	0.1
8					11.2	25.8	24.5	0.3			1.2	0.2
9					24.9	36.5		0.8			2.0	0.2
10					53.9			1.2			2.9	0.6
11								1.8			4.7	0.7
12								2.2			5.5	1.0
13								3.6			9.0	1.3
14								4.4			12.8	1.3
15								4.6			16.7	1.8
16								6.6			22.5	2.1
17								8.6			21.0	2.3
18								11.9				3.2
19								15.2				3.5
20								20.9				4.4
21								16.8				5.9
22												7.1
23												8.7
24												11.4
25												14.8
26												16.6
27												11.9
No. of Items in Sub-Test	5	5	5	5	10	9	8	21	6	6	17	27

[a] There are missing data amounting to 0.3% of students (weighted) or 26 students (unweighted) associated with each Sub-Test and with the Total Test. These students did not attempt any item in the Reading Test but did participate in other parts of the testing program. They do not include the 4 students who attempted the test but failed to obtain any correct answers.

Table A5.2 Distribution of Scores on Numeration Sub-Tests and Total Test for 10-Year-Old Australian Students

No. of Items Correct	Percent of Students (Weighted)[a]															
	Content: Numbers (Whole & Fraction)	Content: Measurement: General	Content: Measurement: Money	Content: Measurement: Time	Ability: Knowledge	Ability: Computation	Ability: Application	Usage: Community	Usage: Classroom	Operation: Addition	Operation: Subtraction	Operation: Multiplication	Operation: Division	Link Items 1975/80	Age Anchor Items	Total Test
0	0.1	0.3	0.9	2.2	0.0	3.9	0.3	0.1	0.1	0.2	1.3	2.3	3.6	0.1	0.2	0.0
1	0.2	1.6	1.1	2.8	0.0	11.1	0.1	0.2	0.1	0.5	4.1	3.5	19.5	0.1	0.4	0.0
2	0.2	6.2	2.6	4.5	0.1	18.3	0.1	0.1	0.2	1.1	8.7	11.4	14.4	0.3	0.6	0.1
3	0.4	13.4	8.6	6.9	0.1	26.0	1.0	0.1	0.3	3.3	16.0	28.5	25.2	0.4	1.5	0.1
4	0.7	24.3	27.1	11.0	0.1	40.5	0.8	0.3	0.6	18.3	26.8	53.7	37.0	0.8	2.6	0.0
5	1.5	35.8	59.4	18.4	0.4		1.5	0.3	1.2	76.4	42.6			1.0	3.5	0.1
6	1.6	18.2		23.7	0.5		2.4	0.6	2.0					2.3	4.4	0.1
7	2.7			30.6	0.6		3.9	0.6	2.9					3.6	6.7	0.1
8	3.6				1.1		5.0	0.7	4.0					5.8	8.5	0.1
9	5.2				1.3		7.3	1.1	3.9					8.3	12.4	0.1
10	7.9				2.5		10.5	1.5	8.0					13.9	17.0	0.4
11	10.2				3.2		15.2	2.4	10.1					21.4	21.0	0.1
12	16.0				4.8		22.3	3.0	15.4					27.9	21.1	0.5
13	20.1				6.7		28.9	3.4	20.6					13.8		0.4
14	29.2				10.3			4.0	29.5							0.4
15					14.4			6.1								0.7
16					19.7			8.1								1.0
17					22.7			11.2								1.0
18					11.2			12.7								1.4
19								17.7								1.2
20								15.9								1.5
21								8.0								1.6
22																2.1
23																2.4
24																2.7
25																3.6
26																3.8
27																5.2
28																5.9
29																7.1
30																8.1
31																9.5
32																10.8
33																11.7
34																10.4
35																5.5
No. of Items in Sub-Test	14	6	5	7	18	4	13	21	14	5	5	4	4	13	12	35

[a] There are missing data amounting to 0.1% of students (weighted) or 7 students (unweighted) associated with each Sub-Test and with the Total Test. These students did not attempt any item in the Numeration Test but did participate in other parts of the testing program. They do not include the 5 students who attempted the test but failed to obtain any correct answers.

Table A5.3 Distribution of Scores on Reading Sub-Tests and Total Test for 14-Year-Old Australian Students

No. of Items Correct	Percent of Students (Weighted)[a]													
	Ability: Comprehending Simple Sentences	Ability: Comprehending Complex Sentences	Ability: Prose Comprehension	Ability: Reference Skills (Classroom)	Ability: Reference Skills (Newspaper)	Ability: Newspaper Comprehension	Form: Words in Context	Form: Reference Materials	Form: Continuous Prose	Usage: Classroom	Usage: Community	Link Items 1975/80	Age Anchor Items	Total Test
0	0.0	0.1	0.1	0.2	0.7	0.3	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.0
1	0.0	0.4	1.6	0.7	1.1	1.0	0.0	0.1	0.1	0.0	0.3	0.0	0.0	0.0
2	0.0	0.7	5.2	1.7	4.2	1.5	0.0	0.1	0.4	0.0	0.4	0.4	0.0	0.0
3	0.1	3.2	9.6	4.5	15.0	4.2	0.0	0.3	0.6	0.0	0.6	1.2	0.0	0.0
4	1.4	14.0	16.9	16.6	37.1	7.0	0.0	0.6	1.4	0.0	1.0	4.3	0.0	0.0
5	98.2	81.5	35.7	76.2	41.6	8.3	0.2	1.1	2.3	0.0	1.4	12.1	0.0	0.0
6			30.7			16.2	0.3	1.8	3.8	0.0	2.3	29.4	0.0	0.0
7						26.6	0.6	2.8	4.6	0.0	4.3	52.4	0.1	0.0
8						34.9	1.9	7.3	5.9	0.1	6.2		0.2	0.0
9							5.9	16.7	8.9	0.0	8.5		0.1	0.0
10							22.4	33.0	12.0	0.2	11.8		0.3	0.0
11							68.5	36.0	19.4	0.3	19.2		0.6	0.2
12									22.9	0.3	23.5		0.9	0.0
13									17.6	0.5	20.2		1.7	0.0
14										1.1			3.8	0.2
15										1.5			8.2	0.2
16										2.0			23.0	0.3
17										4.2			60.9	0.2
18										6.5				0.6
19										10.2				0.4
20										17.5				0.3
21										30.9				0.9
22										24.5				1.1
23														1.2
24														2.1
25														2.1
26														3.1
27														3.8
28														4.2
29														6.5
30														7.9
31														10.0
32														14.2
33														15.3
34														16.0
35														8.7
No. of Items in Sub-Test	5	5	6	5	5	8	11	11	13	22	13	7	17	35

[a] There are missing data amounting to 0.1% of students (weighted) or 9 students (unweighted) associated with each Sub-Test and with the Total Test. These students did not attempt any item in the Reading Test but did participate in other parts of the testing program.

Table A5.4 Distribution of Scores on Numeration Sub-Tests and Total Test for 14-Year-Old Australian Students

No. of Items Correct	Percent of Students Weighted[a]																
	Content: Whole Numbers	Content: Fractions and Decimals	Content: Measurement - General	Content: Measurement - Money	Content: Measurement - Time	Ability: Knowledge	Ability: Computation	Ability: Application	Usage: Community	Usage: Classroom	Operation: Addition	Operation: Subtraction	Operation: Multiplication	Operation: Division	Link Items 1975/80	Age Anchor Items	Total Test
0	0.0	0.4	0.5	0.9	0.3	0.0	0.2	0.2	0.1	0.0	0.1	0.1	0.3	0.9	0.1	0.0	0.0
1	0.1	0.9	0.9	1.3	1.0	0.0	0.3	0.4	0.4	0.0	0.5	0.4	1.5	1.9	0.4	0.0	0.0
2	0.1	3.3	2.3	2.9	2.5	0.2	0.7	0.5	0.4	0.0	3.8	1.1	4.8	4.5	0.6	0.1	0.0
3	0.2	5.6	3.9	8.5	7.4	0.2	1.8	0.5	0.3	0.1	21.8	2.1	12.6	11.6	0.7	0.2	0.0
4	0.4	9.5	8.4	22.1	23.7	0.2	4.5	0.5	0.6	0.0	73.6	4.4	30.9	26.6	1.0	0.1	0.0
5	0.8	16.9	15.3	64.2	65.0	0.5	12.6	0.9	0.8	0.1		7.6	49.7	54.3	1.5	0.3	0.1
6	1.6	26.4	26.7			0.9	29.5	1.1	1.0	0.2		12.4			2.6	0.4	0.1
7	2.0	36.9	41.9			2.3	50.3	1.3	1.2	0.2		25.4			4.1	0.9	0.1
8	6.3					5.1		1.5	1.7	0.4		46.3			6.3	1.4	0.0
9	13.0					12.0		2.5	1.9	0.7					8.6	4.6	0.1
10	27.4					26.1		3.7	2.9	0.8					11.9	9.7	0.2
11	47.9					52.3		4.6	4.9	1.2					15.9	26.8	0.1
12								6.1	6.9	1.9					20.7	55.5	0.1
13								7.6	9.3	3.7					25.6		0.2
14								11.0	14.0	6.1							0.4
15								14.4	22.9	8.5							0.4
16								20.1	30.5	13.0							0.4
17								23.3		17.6							0.5
18										21.9							0.6
19										23.5							0.7
20																	0.8
21																	1.0
22																	1.4
23																	1.3
24																	2.0
25																	3.0
26																	2.8
27																	3.6
28																	5.1
29																	5.4
30																	7.0
31																	9.2
32																	10.4
33																	12.6
34																	15.4
35																	14.7
No. of Items in Sub-Test	11	7	7	5	5	11	7	17	16	19	4	8	5	5	13	12	35

a There are missing data amounting to 0.1% of students (weighted) or 4 students (unweighted) associated with each Sub-Test and with the Total Test. These students did not attempt any item in the Numeration Test but did participate in other parts of the testing program. They do not include the 2 students who attempted the test but failed to obtain any correct answers.

Reading and Numeration

Sub-Test and Test Reliability

Reliabilities for all reading and numeration sub-tests and total tests were estimated using the Marshall Beta Index (Marshall, 1975). The index is equivalent to the mean split-half coefficient of agreement and is analogous to the Cronbach alpha coefficient interpreted as the mean of all possible split-half reliability coefficients (Izard, 1978:316). The Marshall index also tends to be a conservative estimate of test reliability and has the advantage of relative stability under varying sample designs (Izard, 1978:326).

The reading and numeration sub-test and test reliabilities are shown in Table A5.5. Reliabilities tended to be lower for the tests prepared for the 10-year-old students although only in five out of 26 sub-tests did the index fall below 0.80. The reading and numeration tests taken as a whole had reliabilities exceeding 0.80 at the 10-year-old level and exceeding 0.90 at the 14-year-old level. Some of the shorter sub-tests at the 14-year-old level had reliabilities approaching one. As can be observed in the tables in this appendix, these were the very easy sub-tests where almost all students obtained very high scores.

Table A5.5 Marshall Beta Index of Reliability for Reading and Numeration Sub-tests and Total Tests for 10- and 14-Year-Old Students

Sub-test	10-Year-Old Students		14-Year-Old Students	
	No. of Items	Reliability	No. of Items	Reliability
Reading				
Simple Sentences	5	0.95	5	1.00
Complex Sentences	5	0.87	5	0.97
Prose Comprehension	5	0.83	6	0.81
Ref. Skills (Class)	5	0.82	5	0.95
Ref. Skills (News)	-	-	5	0.84
Newspaper Comprehension	-	-	8	0.81
Words in Context	10	0.88	11	0.98
Reference Materials	9	0.81	11	0.93
Continuous Prose	8	0.76	13	0.81
Classroom Usage	21	0.80	22	0.88
Community Usage	6	0.80	13	0.81
Link Items	6	0.80	7	0.86
Age Anchor Items	17	0.80	17	0.96
Total Test	27	0.83	35	0.91
Numeration				
Whole Numbers	-	-	11	0.94
Fractions & Decimals	-	-	7	0.79
Numbers (Whole & Fractions)	14	0.85	-	-
Measurement-General	6	0.74	7	0.81
Measurement-Money	5	0.90	5	0.91
Measurement-Time	7	0.77	5	0.92
Knowledge	18	0.80	11	0.95
Computation	4	0.83	7	0.85
Application	13	0.84	17	0.87
Community Usage	21	0.82	16	0.87
Classroom Usage	14	0.85	19	0.89
Addition	5	0.96	4	0.98
Subtraction	5	0.81	8	0.85
Multiplication	4	0.92	5	0.86
Division	4	0.86	5	0.87
Link Items	13	0.71	13	0.81
Age Anchor Items	12	0.73	12	0.94
Total Test	35	0.84	35	0.90

Writing Tests

Reliability of Global Scores

Three writing tasks at each age level were globally scored on a five-point scale by two markers working independently. Where the scores awarded to any one response differed by more than one point, a third person acted as arbiter and re-scored that answer. The two scores for each answer were then added to provide the total score recorded for each student.

The reliability of these scores has been calculated in terms of product-moment correlation coefficients between the two markers and are reported in Table A5.6. Although the correlations were not as high as might be wished, they were all within the range of 0.76 to 0.79.

Table A5.6 Product-Moment Correlations for Scores Awarded by Two Independent Markers of Each Writing Task Globally Scored

Task	Correlation r	N
10-Year-Old Level		
Write a Report	0.77	2697
Write About Yourself	0.77	2695
Write a Story	0.79	2762
14-Year-Old Level		
Write a Letter	0.79	2534
Write Your Opinion	0.79	2527
Write a Story	0.76	2540

APPENDIX SIX

PROJECT COSTS

The budget for the Australian Studies in Student Performance project was $90,000 in the financial year 1979/80 and $140,000 in 1980/81. The costs of the project are shown in Table A6.1 under five categories of expenditure: namely staff, data processing, travel, printing etc, and accommodation/overheads.

Owing to the project not being commenced until late in 1979, the budget for the year 1979/80 was underspent by approximately $25,000. The late start meant that additional professional staff were employed for part of the year 1980/81 to enable work schedules to be met and the unspent funds from the previous year were utilized. There were other budget changes such as a lower expenditure on clerical/technical staff being balanced by a higher expenditure on data processing. Printing costs were higher than expected, in part because increases in printing costs had been underestimated, and in part because the different types of items in some of the tests meant that these tests had many more pages than those originally costed.

It was considered that the cost of the program per student tested would be of interest and this is also shown in Table A6.1. The overall cost per student was $22.27, made up of staff costs of $15.88 (71 per cent) and other expenses $6.40 (29 per cent). In general it would be found that for a project of this size and nature staff costs would account for something more than 70 per cent of the total cost. However if the total sample size of approximately 10,500 students were varied considerably these proportional costs would also change.

Table A6.1 Project Costs 1979-1981

Item	1979/80 $	1980/81 $	Total $	Cost per Student($)
1. Staff				
a. Professional	44,259	88,665	132,924	12.57
b. Part-time clerical	2,121	20,110	22,231	2.10
c. Casual (Prof. & Cler.)	-	1,983	1,983	0.18
d. Secretarial (typing)	4,218	6,560	10,778	1.02
2. Data Processing	749	16,001	16,750	1.58
3. Travel	927	748	1,675	0.16
4. Printing, stationery, duplicating, postage, telephone	5,922	26,430	32,352	3.06
5. Accommodation/overheads	6,850	10,000	16,850	1.59
Total	65,046	170,497	235,543	22.27